

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 186 a 246 c 224 g 178 t
ORIGIN

Query Match 28.2%; Score 493; DB 10; Length 834;
Best Local Similarity 100.0%; Pred. No. 3e-228;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

664 gctgggtcgctgaggttgactgggacagctctctgcttaaaatctactctggtcctc 723
Db 1 GCTGGGCTGCTGAGTTGACTGGGACAGCTCTGCTTAAATCTACTCTGGGCTCTC 60

724 ccatacagtggttccatctcttagcagcaactgaatgaactcctactcactcagagacctg 783
Db 61 CCATCAGTGGGTCCCACTCTAGCAGCAACTGGAATGACTCTTACTCAGAGAGACCTG 120

784 ccagcagctgggttctgagagtgctccaccgacacacagagtggtgcccacaggtttgc 843
Db 121 CCAGCAGTGGGTTCGAGAGTGTCTACCGGACACACGAGGTGCCCCACAGGATTTGC 180

844 caacagcttctcaatctttagatatacaactccaccatccaggaagcctccacaggtctga 903
Db 181 CAACAGCTTCTCAATCTTGGAGATACAACTCCACCATCCAGGAAAGCTCCACAGTCTGA 240

904 atgcccctccaggggtatctccctccagtggttccactgctggagactgagggccatgac 963
Db 241 ATGCCCTTCCAGCGGTATATCTCCCTCCAGTGTCCCACTGCGGACTGAGGGCCATGAC 300

964 cggcgagctgctggagggcgctggctcgatagcaagtggccttgccaagtgaactc 1023
Db 301 CGGCGGATCTGGAGGGGGCTGGCTCGGATAGCAAGTGGCTTGGCAAGTGAAGTCT 360

1024 gcaattcggcaccacccacatctgtggaggcagcgtcattgacgccagtggtggtctcac 1083
Db 361 GCACCTTGGCACCACCCACATCTGTGGAGGACGCGTCTATTGACGCCAGTGGTCTCAC 420

1084 tcccccactgcttcttgacccgggagagtgctgagggctgagggctgagaggtgtaacg 1143
Db 421 TGGCGCCACACGCTCTCTGCTGGAGCCCGGAGAGGCTCTGGAGGGCTGGAAGGTGTAACG 480

1144 gggcaccagcaac 1156
Db 481 GGGCACCAGCAAC 493

RESULT 2
AW845106 285 bp mRNA linear EST 19-MAY-2000
LOCUS
DEFINITION M80-CT0006-010699-022 CT0006 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW845106
VERSION AW845106.1 GI:7940623
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 285)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-MR0-CT0006-010699-022&t3=1999-06-01&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 285.

FEATURES
source

1..285
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0006"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site: 1; SmaI; Site: 2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 1,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 59 a 85 c 90 g 51 t
ORIGIN

Query Match 16.3%; Score 285; DB 9; Length 285;
Best Local Similarity 100.0%; Pred. No. 2.5e-127;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 957 ccatgaccggcgatcgctggagggcgctgctcgatgtagcaagtggccttggaag 1016
Db 1 CCATGACCGGGCGGATCGTGGAGGGGCTGGCTCGGATAGCAAGTGGCTTGGCAAG 60

Qy 1017 tgagtctgacttcggcaccacacacatctgtgagcagctattgacccagtgagg 1076
Db 61 TGAGTCTGACTTCGGGACCCACCATCTGTGGAGGACGCTCATTCACGCCAGTGGG 120

Qy 1077 tgctcactgcgcgcactgcttcttcgtgacccgggagaggtcctcgagggctggaag 1136
Db 121 TGCTCACTCGCGCCACTGCTTCTTCGTGACCCGGAGAGGCTCTGGAGGCTGGAAGG 180

Qy 1137 tgtaaggggcaccagcaaacctgacacagttgctgtagggagcctcactgcagatca 1196
Db 181 TGTACGGGGCACCAGCAACCTGCACAGTTGCTGAGGACGCTCCATTGCCGAGATCA 240

Qy 1197 tcatacagcaattaccacgagtaggagagcactatgacatcg 1241
Db 241 TCATCACAGCAATTACCCGATGAGGAGGAGGAGGACTATGACATCG 285

RESULT 3
BF757612 252 bp mRNA linear EST 12-JAN-2001
LOCUS
DEFINITION I10-CT3002-071100-500-a05 CT3002 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF757612
VERSION BF757612.1 GI:12105616
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 252)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Other ESTs: zt19a12.x5
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information
This 5' resequenced clone has no previous 5' data to verify this
new read against
Insert Length: 739 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 467.
Location/Qualifiers
1..591
/organism="Homo sapiens"
/db_xref="GDB:593387"
/db_xref="taxon:9606"
/clone="IMAGE:713566"
/clone.lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo."

FEATURES
sourceBASE COUNT
ORIGIN

161 a 160 c 126 g 142 t 2 others

Query Match 8.8%; Score 154; DB 9; Length 591;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 gagggctggaaggtgtacggcgccgagcagcaacctgacacgttgctgagcgagcctcc 1183
|||||
Db 1 GAGGCTGGAAGGTGTACGGCGCCGACGACCACTGCACCACTGCTGAGCGAGCCTCC 60
|||||

QY 1184 attccagatcatcaacagcaattacacccatgagggagcgactatgacatgcc 1243
|||||
Db 61 ATTCCGAGATCATCAACAGCAATTACACCGATGAGGAGCGAGCGACTATGACATGCC 120
|||||

QY 1244 ctcatggcgtgtccaaagccctgacctgtccg 1277
|||||
Db 121 CTCATGGCGTGTCCAGGCCCTGACCTGTCCG 154
|||||

RESULT 8

AI820925 600 bp mRNA linear EST 09-JUL-1999
LOCUS
DEFINITION
AI820925
2053f08.y5 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:741735 5' similar to contains TAR1.c1 TAR1 repetitive element
;; mRNA sequence.

ACCESSION

AI820925
AI820925.1 GI:5440004

VERSION

EST.

KEYWORDS

human.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 600)

AUTHORS

NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)

COMMENT

Other ESTs: zu53f08.s1
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: WashU-NCI human EST Project
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 461.
Location/Qualifiers
1..600
/organism="Homo sapiens"
/db_xref="GDB:5941928"
/db_xref="taxon:9606"
/clone="IMAGE:741735"
/clone.lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo."

FEATURES
sourceBASE COUNT
ORIGIN

164 a 162 c 128 g 145 t 1 others

Query Match 8.8%; Score 154; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 gagggctggaaggtgtacggcgccgagcagcaacctgacacgttgctgagcgagcctcc 1183
|||||
Db 1 GAGGCTGGAAGGTGTACGGCGCCGACGACCACTGCACCACTGCTGAGCGAGCCTCC 60
|||||

QY 1184 attccagatcatcaacagcaattacacccatgagggagcgactatgacatgcc 1243
|||||
Db 61 ATTCCGAGATCATCAACAGCAATTACACCGATGAGGAGCGAGCGACTATGACATGCC 120
|||||

QY 1244 ctcatggcgtgtccaaagccctgacctgtccg 1277
|||||
Db 121 CTCATGGCGTGTCCAGGCCCTGACCTGTCCG 154
|||||

RESULT 9

R78581 402 bp mRNA linear EST 09-JUN-1995
LOCUS
DEFINITION
Y173c10.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:144882 3' similar to SP:HEPS.RAT Q05511 SERINE PROTEASE
HEPS1 : contains MER22 repetitive element ; mRNA sequence.

ACCESSION
VERSION
R78581
R78581.1 GI:854862
KEYWORDS
EST.
SOURCE
human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 402)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston,
R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK
Washington University School of Medicine

Query Match		25.9%; Score 777.5; DB 1; Length 490;
Best Local Similarity		32.7%; Pred. No. 4e-35;
Matches	180; Conservative	81; Mismatches 201; Indels 89; Gaps 15;
OY	27	ACPPGGRASPAQASPAQAGTTPPGRASPAQASPACTPPGRASPGRASPAQASPARASP 86
Db	5	SGSPGG-IGPCYNHCYQSEHICPP-----RPPVAPNGYNLYPQAYYYP----- 46
OY	87	ALASLSRSSGSRSSARSASVTPTRYLVLRATPVCAVP-IRSSSPARS-APATRATRES 144
Db	47	-----SPVPQYAPRTTQASTSVIHTHPKSSGAPCTSKSKS 83
OY	145	PQTSLPKFTWREGOKOLPLIGCVLLLTIALVSTLIIFQFMQGHGTIRYKQRES--CPKH 202
Db	84	LCLIALAIGT-----VLITGAAVAALLWRFWDSNCSTSEMESCSSGTCISS 128
OY	203	AVRCDGVVDCKLSDGLCVRFWDWSLLAIYSGSSHQWLPICSSNNWDSISEKTQQQLG 262
Db	129	SUWCQGVHAHCPNGEDENRCRVLYQGOSFILLQVYSQRKAWAYPVQCDDWSESYGRAACKDMG 188
OY	263	FESAHRTE-VAHROPANGFSILRYNS---TIQESLHRSECPSQRYISLCQSHCGLRAM 317
Db	189	YKNFNYSOGIPDQSCATCFMKLNWSSGNVDLYKKYLHSDCSRRVMVLSURCECGVRSV 248

QY 1321 tagctcaatgagacagctgtgatcacaggttttggcaagaccagggagacagatgacaa 1380
 |||||
 Db 1398 TAGCCTCAATGAGACCTGTGTGATCACAGGCTTTGGCAAGACAGGAGACAGATGACAA 1457
 |||||
 QY 1381 gacatcccccttccctccggaggtgacaggtcaatctcatcacttcaagaatgaatga 1440
 |||||
 Db 1458 GACATCCCCCTTCCCTCCGGAGGTGAGGTCAATCTCATGCACTTCAAGAATGAATGA 1517
 |||||
 QY 1441 ctacttggtctatgacagttaccttaccaccaaggtatgtgtgctggggaccttcgtgg 1500
 |||||
 Db 1518 CTACTTGGCTATGACAGTTACTTACCCCAAGGATGATGTGTGGGGACCTTCGTGG 1577
 |||||
 QY 1501 gggcaagactcctccaggggagacagggggggccttctgtgtgagcagaacacgg 1560
 |||||
 Db 1578 GGCACAGAGACTCTGTCAGAGGAGACAGGGGGGGGCTCTTGTGTGAGCAGAAACACG 1637
 |||||
 QY 1561 ctggtacctgacaggtgtccacagctgggacagctgtggtccagagaaacaaacctgg 1620
 |||||
 Db 1638 CTGCTACCTGGCAGGTGTGACAGCTGGGGACAGGCTGTGGCCAGAGAAACAACTGG 1697
 |||||
 QY 1621 tgtgtacacaaagtacagaagttcttccctggatttacagcaagatggag 1672
 |||||
 Db 1698 TGTGTACACCAAGTCACAGAAGTTCTTCCCTGGATTTACAGCAAGATGGAG 1749
 |||||

RESULT 3
 AXI49581
 LOCUS AXI49581 2067 bp DNA linear PAT 08-JUN-2001
 DEFINITION Sequence 5 from Patent WO0136604.
 ACCESSION AXI49581
 VERSION AXI49581.1 GI:14348015
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2067)
 AUTHORS Madison,E.L. and Ong,E.O.
 TITLE Nucleic acids encoding endothelialases, endothelialases and uses
 thereof
 JOURNAL Patent: WO 0136604-A 5 25-MAY-2001;
 CORVAS INTERNATIONAL, INC. (US)
 FEATURES
 Location/Qualifiers
 1..2067
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2067
 /note="DNA sequence encoding a transmembrane serine
 protease (endothelase-L) protein"
 /codon_start=1
 /protein_id="CAC41221.1"
 /db_xref="GI:14348016"
 /translation="MERDSDHGNASPARTPSAGASPAQASPAQSPAGTPPGRASPAQASPAQA
 SPACTPPRASPAQASPAQTPPGRASPAQASPAQSPALASLSRSSGRSSA
 RSASVTPSPRVLRATPVCAVPIRSPARSAPATRESPTSLPFTWREGNQ
 LPLGVLDLILVLSLILFQWGHGIRYKEQRECPKHAQVRCQVYDCKLSD
 LGVRFQWDLKSLIKYSGSHQWLPICSSNNDSYSKTCQQLGFESAHRTTEVAHRD
 FANSFSLRYNLTQESLHRSQPSQRYISLQSHCGLRAMTGRIVGALASDKWP
 QVSLHFGTTHICGGLTDAQWLTAAHCFVTRKYLEGKVKVAGTSLNHLPEASAI
 AEIINSNYDDEDDYDALMRLSKPLTSLAHHPACLPMHGOTFSLNETCWITGPK
 TREDDTSPFLREVQNLIDFKCNDYLVITSLYTPRMCAQDLRGRDSCQDSSG
 PLVCEQNNRLVAGTWSWGGQGNKPKYVTVKTEVLPWYIKSMENRAQRVKAFTY
 RPKQLGRCSPRSIFLKVAMPDFVSVSAEDFVIFVILKLMGRIRSSWPPALFV
 LVFLIFFLLLLSLFKNTSDSILTLTFTAVTRMLPENYHSFPFLKIPHLQNTTIK
 EYKVIK"
 BASE COUNT 463 a 604 c 540 g 460 t
 ORIGIN

Query Match 95.2%; Score 1663.4; DB 6; Length 2067;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1664; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 20:45:23 ; Search time 1890.31 seconds
(without alignments) 12480.851 Million cell updates/sec

Title: US-09-879-792-11
Perfect score: 1748
Sequence: 1 ctacagaccatgagagagg.....ggctgctgtaactcgagaaa 1748

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_estl.*
10: gb_est2.*
11: gb_htc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	28.2	834	10 BE732381	BE732381 601569601
2	285	16.3	285	9 AW845106	AW845106 MR0-CT000
3	239	13.7	252	10 BF757612	BF757612 IL0-CT300
4	239	13.7	254	10 BF757611	BF757611 IL0-CT300
5	239	13.7	255	10 BF757610	BF757610 IL0-CT300
6	185	10.6	190	9 BE073527	BE073527 RC5-BT055
7	154	8.8	591	9 AI820898	AI820898 zt19a12.y
8	154	8.8	600	9 AI820925	AI820925 zu53f08.y
9	151	8.6	402	10 R78581	R78581 y173c10.r1
10	139	8.0	286	9 AI909842	AI909842 QV-BT225-
11	134	7.7	751	10 BE280394	BE280394 601158674
12	103	5.9	235	9 AA402094	AA402094 zu53f08.f
13	101	5.8	384	9 AA285124	AA285124 zt23e06.r
14	52	3.0	471	10 R78582	R78582 y173c10.sl
15	49	2.8	614	10 BE286322	BE286322 601095164
16	49	2.8	618	12 BH098720	BH098720 RPCI-24-2
17	49	2.8	622	10 BF119003	BF119003 601755367

18	38	2.2	621	10 BE285038	BE285038 601098725
19	38	2.2	678	10 BE289529	BE289529 601088060
20	29	1.7	563	10 BE290088	BE290088 601089056
21	24	1.4	317	10 BF750399	BF750399 RC1-BN041
22	24	1.4	523	9 AA542994	AA542994 ni55b01.s
23	24	1.4	527	9 AA101043	AA101043 zm27ell.r
24	24	1.4	619	10 W58737	W58737 zd23607.r1
25	24	1.4	711	9 AU134435	AU134435 AU134435
26	24	1.4	914	10 BF601676	BF601676 603249916
27	24	1.4	1169	10 BG677775	BG677775 602625451
28	23	1.3	688	10 BE309103	BE309103 601093231
29	22	1.3	940	10 BE693852	BE693852 603342491
30	21	1.2	130	9 BE075424	BE075424 MK2-BT058
31	21	1.2	237	9 AA861735	AA861735 ak38b11.s
32	21	1.2	299	9 AA397523	AA397523 zt72c07.s
33	21	1.2	336	9 AW326772	AW326772 19750 MAR
34	21	1.2	415	10 BE909673	BE909673 601501682
35	21	1.2	473	9 AW659470	AW659470 96665 MAR
36	21	1.2	480	9 A1138859	A1138859 qe04d10.x
37	21	1.2	482	9 A1138858	A1138858 qe04d09.x
38	21	1.2	558	9 AA393577	AA393577 zt72c07.f
39	21	1.2	575	12 A2227042	A2227042 RPCI-23-6
40	21	1.2	655	12 A2801541	A2801541 2M0060305
41	21	1.2	752	10 BE827673	BE827673 603074273
42	21	1.2	817	10 BF139806	BF139806 601785786
43	21	1.2	904	10 BG823531	BG823531 602728910
44	21	1.2	906	10 BF691078	BF691078 602247133
45	20	1.1	186	9 BB182766	BB182766 BB182766

ALIGNMENTS

RESULT 1
BE732381 834 bp mRNA linear EST 15-SEP-2000
LOCUS 601569601F1 NTH_MGC_21 Homo sapiens cDNA clone IMAGE:3844129 5',
DEFINITION mRNA sequence.
ACCESSION BE732381
VERSION BE732381.1 GI:10146373
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CW540 row: b column: 02
High quality sequence stop: 817.
location/Qualifiers
1. 834
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3844129"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="Dhl0B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

BASE COUNT      186 a      246 c      224 g      178 t
ORIGIN

Query Match      28.2%; Score 493; DB 10; Length 834;
Best Local Similarity 100.0%; Pred. No. 3e-228;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 664 gctgggtcggtgagttgactgggacaaagtctctgtcttaaaatctactctgggtctc 723
Db 1 GCTGGGTCGCGTAGGTTGACTGGGCAAGTCTCTGTTAAAATCTACTCTGGGTCTC 60

Qy 724 ccactagtggttcccatctgtacagacaactgaatgactcctactcagagaagactg 783
Db 61 CCATCAGTGGGTTCCCACTCTAGCAGCACTGGAACTGACTCTCTCTCAGAGAAGACTG 120

Qy 784 ccagcagctgggtttcagagtgctccaccgacacacagaggttgcacacagggattttgc 843
Db 121 CCAGCAGCTGGGTTTCGAGAGTGTCTACCGGACAAACGAGGTTGCCACAGGGATTTCG 180

Qy 844 caacagcttccaattcttgagatacaactccaccatccaggaagacccacaggtctga 903
Db 181 CAACAGCTTCTCAATCTTTGAGATACAACTCCACCATCCAGAAAAGCCTCCACAGGTCTGA 240

Qy 904 atgcccctccagcgggtatatctccctccagtgctcccaactgcggactgagggccatgac 963
Db 241 ATGCCCTTCCACAGGGATATCTCCTCCAGTGTCCCACTGGCGAGTGGAGGCCATGAC 300

Qy 964 cgggcggatctggtgaggggctggtcctcggtatgacgaagtggccttggcgaagtgtct 1023
Db 301 CGGCGGATCTGTGGAGGGCGCTGGCCTCGGATAGCAAGTGGCTTGGCAAGTGAGTCT 360

Qy 1024 gcaactcggaccaccaacatctgtgagggcacactcattgaacccagtggtgctcac 1083
Db 361 GCACCTCGGACACCCACACATCTGTGGAGGACCGCTCATTTACGCCCCAGTGGGTGCTCAC 420

Qy 1084 tqcccccactgctcttcgtgacccgggagaaagtcctgagggctggaaggtgtacgc 1143
Db 421 TGCCGCCCACTGCTTCTTCGTGACCCGGGAGAGGTCCTCGAGGGCTGGAAGGTGTACGC 480

Qy 1144 gggccaccagcaac 1156
Db 481 GGGCACCAAGCAAC 493

RESULT 2
AW845106 AW845106 285 bp mRNA linear EST 19-MAY-2000
LOCUS MRO-CT0006-010699-022 CT0006 Homo sapiens cDNA, mRNA sequence.
DEFINITION MRO-CT0006-010699-022 CT0006 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW845106
VERSION AW845106.1 GI:7940623
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 285)
AUTHORS Dias Neto,E., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Nagai,M.A., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ft2=MRO-CT0006-010699-022&tl=1999-06-01&st4=1)
Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 285.
FEATURES
Location/Qualifiers
1..285
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0006"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (J.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      59 a      85 c      90 g      51 t
ORIGIN

Query Match      16.3%; Score 285; DB 9; Length 285;
Best Local Similarity 100.0%; Pred. No. 2.5e-127;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 957 ccattgaccggcggtatcggtggagggcgctggcctcgatagcaagtggccttggcgaag 1016
Db 1 CCATGACGGGCGGATCGTGGAGGGCGCTGGCTCGGATAGCAAGTGGCTTGGCAAG 60

Qy 1017 tgaacttcacactcggcaccacccacacatctgtggagcgactcttgacgcccagtggg 1076
Db 61 TGAATCTGACACTTCGGGACCCACCCACATCTGTGGAGGCAGCTCATTTGACGCCCAAGTGGG 120

Qy 1077 tgcctcactgcgcgcactgtcttcttgaccgggagaggtcctggaaggctggaagg 1136
Db 121 TGCTCACTGCGCGCCACTGCTTCTTCGTGACCCGGAGAGAGTCTCTGGAGGGCTTGGGAAG 180

Qy 1137 tgatcggggcaccagcaaacctgacacagttgcttcgagcagcctccatgcccagatca 1196
Db 181 TGTACGGGGCACCAGCAACCTGCACCACTGTCCCTGAGGACGCTCCATTCGCCAGATCA 240

Qy 1197 tcatacacagcaattacaccgatgaggagggagcactatgacatcg 1241
Db 241 TCATCACAGCAATATACACCATGAGGAGGAGGAGGACTATGACATCG 285

RESULT 3
BF757612 BF757612 252 bp mRNA linear EST 12-JAN-2001
LOCUS IL0-CT3002-071100-500-a05 CT3002 Homo sapiens cDNA, mRNA sequence.
DEFINITION IL0-CT3002-071100-500-a05 CT3002 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF757612
VERSION BF757612.1 GI:12105616
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 252)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
```


JOURNAL
MEDLINE
COMMENT

PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL06t2-IL0-CT3002-071100-500-a05t3-2000-11-07&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 251.

[illegible]

```

1. 252
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Ct3002"
/dev_stage="Adult"
/note="Organ: colon; Vector:
Smal; A mini-library was made
from ORESTES PCR (U.S. Let
716 - Ludwig Institute fo
into the pUC 18 vector. Re
mRNA and cDNA amplification
stringency conditions."
71 c 74 g 57 t
a

```

BASE COUNT	50 a	71 c	74 g	57 t
ORIGIN	stringency conditions.			

Query Match 13.7%; Score 239; DB 10; Length 252;
Best Local Similarity 100.0%; Pred. No. 5.2e-105;
Matches 239; Conservative 0; Mismatches 0; Indels

514	QY	ctctctcattgcctggtggttttcgtcatactctcttcagttcttgccagggccacac	573
	Db		
1	Db	CTCTCTATTGCCCTGGTGGTTTGGCTTCATCATCTCTTCAGTTCTGGCAGGGCCACAC	60
574	QY	agggatcaggtacaaagacagaggagagactgtccaaacacgccttctgcttgacgg	633
	Db		
61	Db	AGGATTCAGTACAAAGACAGAGGGAGAGCTGTCCCAAGCACCGCTGTGCTGTGACGG	120
634	QY	ggtgtgtgactgcaagctgaagtagtgacagactgggctgcgtgaggtttgactgggacaa	693
	Db		
121	Db	GGTGTGTGACTGCAAGCTGAAGAGTGCAGAGTGGGCTGGCTGAGGTTTGCATGGGACAA	180
694	QY	gtctctgtctaaatctactctctgggtctcccactcagtggtctcccactctgtacgcaa	752
	Db		
181	Db	GTCTCTGCTTTAAATCTACTCTGGGTCTCCCATCATCTGAGTGGCTTCCCATCTGTACAGCA	239

RESULT 4

BF757611	BF757611	254 bp	mRNA	linear	EST 12-JAN-2001
LOCUS	IL0-CT3002-071100-500-a03	Ct3002	Homo sapiens	cdna	mRNA sequence.
DEFINITION	BF757611				
ACCESSION	BF757611.1	GI:12105615			
VERSION					
KEYWORDS	EST.				
SOURCE	human.				

ORGANISM

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates: Catarrhini: Hominoide: Homo. 1 (bases 1 to 254)

Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

REFERENCE
AUTHORS

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/srfs/0500-guthcm12.p1?tl=IL0&t2=IL0-CR3002-0711000-a03&t3=2000-11-07&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 253.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

location/Qualifiers
1. 254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT3002"
/dev_stage="Adult"
/note="Organ: colon; Vector
      Smal; A mini-library was made
      from ORESTES PCR (U.S. Let
      716 - Ludwig Institute fo
      into the pUC 18 vector. Re
      mRNA and cDNA amplification
      stringency conditions."
71 c 75 q 57 t
a

```

BASE COUNT	51 a	71 c	75 g	57 t
ORIGIN	stringency conditions.			

Query Match 13.7%; Score 239; DB 10; Length 254;
Best Local Similarity 100.0%; pred. No. 5.3e-105.

514	QY	ctctctcattgcactggctggttctgctcattcctctcttccagttcttggcagggccacac	573
1	Db	CTTCCTCATTTGCCCTGGTGTGTTTGGCTCATCTCCTTCCAGTTCTGGCAGGCCACAC	60
574	QY	agggatcaggtacaaagacagagggagagctgcccaagcacactcttcgctgtgacgg	633
61	Db	AGGGATCAGGTACAAAGAGCAGAGGGAGAGCTGTCCCAAGCACCTGTTCTGTGACGG	120
634	QY	ggtggtggactgcaagtgaagctgacagagctgggctgctgctgaggtttgactgggacaa	693
121	Db	GTTGGTGGACTGCAAGCTGAAGAGTGACGAGCTGGGCTGGCTGAGGTTTGACTGGGACAA	180
694	QY	gtctctgtctaaatactctctgtgggtcctcccatcagtgcttcccactctgtagcagca	752
181	Db	GTCTCTGCTTTAAATCTACTCTGGGTCTCCCATCACTGGGTTCCCATCTGTACAGCA	239

RESULT 5

BF757610	BF757610	255 bp	mRNA	linear	EST 12-JAN-2001
LOCUS	IL0-CT3002-071100-500-a01	Cr3002	Homo sapiens	cDNA,	mRNA sequence.
DEFINITION	BF757610				
ACCESSION	BF757610.1	GI:12105614			
VERSION					
KEYWORDS	EST.				
SOURCE	human.				

ORGANISM	Homo sapiens
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

REFERENCE .
AUTHORS
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Hominoidei: Homo.
1 (bases 1 to 255)
Dias Neto, E.F., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsumura, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL0&t2=IL0-CT3002-
071100-500-g01&t3=2000-11-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 254.

FEATURES

source
1. .255
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT3002"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 52 a 71 c 75 g 57 t
ORIGIN

Query Match 13.7%; Score 239; DB 10; Length 255;
Best Local Similarity 100.0%; Pred. No. 5.3e-105;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 514 cctcctcattgcccgtggttccgctcattcctcctcagttctggcaggccacac 573
DB 1 CCTCCTCATGCCCCGTGGTTCGCTCATCATCTCCTTCCAGTTCGGCAGGCCACAC 60
QY 574 agggatcagggtacaggagcagagggagagctgtcccaagacagctgttcgctgtgacgg 633
DB 61 AGGGATCAGGTACAGGAGCAGAGGAGAGAGCTGTCCCAAGCACGCTGTTCGCTGTGACGG 120
QY 634 gttggtgactgcaagctgaagtagcagcagctggctggctggctgaggttgactgggacaa 693
DB 121 GGTGGTGGACTGCAAGCTGAAGTAGAGTAGAGCTGGGCTGGCTGAGGTTGACTGGGACAA 180
QY 694 gtctctgtctaaatctactctggtctcccatcagtggtctcccatctctgttagcagca 752
DB 181 GTCTCTGCTTAAATCTACTCTGGGTCTCTCCCATCAGTGGCTCCCATCTGTAGCAGCA 239

RESULT 6

BE073527
LOCUS BE073527 190 bp mRNA linear EST 09-JUN-2000
DEFINITION RC5-BT0559-140200-012-H07 BT0559 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE073527
VERSION BE073527.1 GI:8420588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 190)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=5&t2=RC5-BT0559-140
200-012-H07&t3=2000-02-14&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 190.

FEATURES

source
1. .190
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0559"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 49 a 56 c 41 g 44 t
ORIGIN

Query Match 10.6%; Score 185; DB 9; Length 190;
Best Local Similarity 100.0%; Pred. No. 8.2e-79;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 689 gacaagtctgtctaaatctactctggtctcccatcagtggtctcccatctgtagc 748
DB 1 GACAAGTCTCTGCTTAAATCTACTCTGGTCTCTCCCATCAGTGGCTTCCCATCTGTAGC 60
QY 749 agcaatggaatgactctactcactcagagaagacctgccagcagctgggttcagagagtgc 808
DB 61 AGCAACTGGAATGACTCCTACTCAGAGAAGACCTGCCAGCAGCTGGGTTCGAGAGTGCT 120
QY 809 caccggacaaccgaggttgcacacagggatttgcacagcttctcaatcttgagatac 868
DB 121 CACCGGACAACCGAGGTGGCCACAGGGATTTTGCCAACAGCTTCTCAATCTTGAGATAC 180
QY 869 aactc 873
DB 181 AACTC 185

RESULT 7

AI820898
LOCUS AI820898 591 bp mRNA linear EST 24-OCT-2000
DEFINITION z19a12.y5 Soares ovary tumor N6HOT Homo sapiens cDNA clone
IMAGE:713566 5', similar to contains TAK1.L3 TAR1 repetitive element
, mRNA sequence.

ACCESSION AI820898
VERSION AI820898.1 GI:5439977
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 591)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tissue Procurement: AFCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L1CMI81 row: d column: 20
 High quality sequence start: 6
 High quality sequence stop: 513.
 Location/Qualifiers

FEATURES

source

1. .751
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3505123"
 /clone_lib="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life technologies)."
 BASE COUNT 182 a 221 c 205 g 143 t

BASE COUNT

ORIGIN

Query Match 7.7%; Score 134; DB 10; Length 751;
 Best Local Similarity 100.0%; Pred. No. 7.1e-54;
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 299 ggcaggtcatccatccgaggtcagctcggtgacaccccccacacaggtgacctt 358
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 171 GGCAGTCTATCATCCGCCAGTCTGCTCGGTGACACCTCCGCCACGAGTGTACCTT 230
 QY 359 gtttagagcaacacagtggtgggggtaccatccgatctctctgccaggtcagcacca 418
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 231 GTTAGAGCAACACAGTGGGGGCTGTACCCATCCGATCATCTCTGCCAGTCCAGCACCA 290
 QY 419 gcaaccaggggccac 432
 ||||||||||||||||
 Db 291 GCAACCAAGGGGCCAC 304

RESULT 12

AA402094

LOCUS

DEFINITION AA402094 235 bp mRNA linear EST 16-MAY-1997
 IMAGE:741735 5', mRNA sequence.

ACCESSION

AA402094

VERSION

AA402094.1 GI:2056077

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 235)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
 T., Waterston, R., and Willson, R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 Contact: Willson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

TITLE

JOURNAL

COMMENT

source

FEATURES

source

Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 201.
 Location/Qualifiers
 1. .235
 /organism="Homo sapiens"
 /db_xref="GDB:5941928"
 /db_xref="taxon:9606"
 /clone="IMAGE:741735"
 /clone_lib="Soares ovary tumor NBHOT"
 /sex="Female"
 /tissue_type="ovarian tumor"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: ovary; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGTGGAGCGCGGTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldo."
 BASE COUNT 48 a 88 c 56 g 43 t

BASE COUNT

ORIGIN

Query Match 5.9%; Score 103; DB 9; Length 235;
 Best Local Similarity 100.0%; Pred. No. 5.9e-39;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1175 gcagctccattgcagatcatcatcaacagcaattacaccgagggagcgactat 1234
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 49 GCAGCTCCATTCGCCGAGATCATCATCAGCAGATATACACCGAGGAGGAGCATAT 108
 QY 1235 gacatgcctcatgcgctgtccaaagccctgacccgtgccg 1277
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 109 GACATGCCCTCATCGCTGTCTCAAGCCCTCGACCTGTCCG 151

RESULT 13

AA285124

LOCUS

DEFINITION AA285124 384 bp mRNA linear EST 15-MAY-1997
 IMAGE:713986 5', mRNA sequence.

ACCESSION

AA285124

VERSION

AA285124.1 GI:1927878

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 384)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
 T., Waterston, R., and Willson, R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 Contact: Willson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28ml3 rev2 ET from Amersham.
 Location/Qualifiers
 1. .384
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:713986"
 /clone_lib="Soares ovary tumor NBHOT"
 /sex="Female"

FEATURES

source

```

/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAATGGAGCGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
BASE COUNT      100 a      120 c      84 g      80 t
ORIGIN

```

```

Query Match      5.8%; Score 101; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 6.3e-38;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1177 agctccatgccgagatcatcaacagcaattacaccgatgagagcagactatga 1236
Db 53 AGCCTCATTCGCGAGATCATCATCAACAGCAATTACACCGATGAGGAGCGACTATGA 112
Qy 1237 catcgccctcatgctgtccaaagccctgagccctgtccg 1277
Db 113 CATGCCCTCATGCGGTGTCCAAAGCCCTGACCCCTGTCCG 153

```

```

RESULT 14
R78582/c
LOCUS      R78582      471 bp      mRNA      linear      EST 04-AUG-1997
DEFINITION y173c10.s1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:144882 3' similar to SP:KAL_MOUSE P26262 PLASMA KALLIKREIN
PRECURSORS ; mRNA sequence.
ACCESSION R78582.1 GI:854863
VERSION R78582
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 471)
Hillier, J., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoso, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins
, M., Hultman, M., Kucaba, R., Lacy, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert Size: 554
High quality sequence stops: 153
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Insert Length: 554 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 153.

```

```

FEATURES
source
1..471
/organism="Homo sapiens"
/db_xref="GDB:558601"
/db_xref="taxon:9606"
/clone="IMAGE:144882"
/clone_lib="Soares placenta Nb2HP"

```

```

/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
ACTGGAAGAATTCGCGCGCGAGAAATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      93 a      105 c      133 g      118 t      22 others
ORIGIN

```

```

Query Match      3.0%; Score 52; DB 10; Length 471;
Best Local Similarity 100.0%; Pred. No. 4.2e-14;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1209 attacaccgatgagagcagactatgacatgcctcatgctgtccaa 1260
Db 298 ATTACCCGATGAGGAGGAGACTATGACATGCCCTCATGCGGCTGTCCAA 247

```

```

RESULT 15
BE286322
LOCUS      BE286322      614 bp      mRNA      linear      EST 26-OCT-2000
DEFINITION BE286322.1 GI:9164563
ACCESSION BE286322
VERSION BE286322.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-f@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8531 row: o column: 15
High quality sequence stop: 569.

```

```

FEATURES
source
1..614
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:349002"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

```

```

BASE COUNT      159 a      168 c      161 g      126 t
ORIGIN

```

```

Query Match      2.8%; Score 49; DB 10; Length 614;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;

```

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1559 cgctggtacctggcaggtgtcaccagctggggcacaggtgtggccaga 1607
 |||||
Db 369 CGCTGGTACCTGGCAGGTGTCACCAGCTGGGGCACAGGCTGTGGCCAGA 417

Search completed: August 19, 2002, 22:54:10
Job time: 7727 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:42:24 ; Search time 88.44 Seconds
(without alignments)
1099.312 Million cell updates/sec

Title: US-09-879-792-12

Perfect score: 2999

Sequence: 1 MERDSHGNSPARTPSAGAS.....TEVLPTWYKMESEVRFRKS 562

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL_19.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mhc.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phase.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2942	98.1	581	4 Q9BYE2	Q9bye2 homo sapien
2	2761.5	92.1	537	4 Q9BYE1	Q9bye1 homo sapien
3	2207	73.6	491	4 Q96YI8	Q96iy8 homo sapien
4	325	30.8	176	11 Q91VQ8	Q91vq8 mus musculus
5	763.5	25.5	490	11 Q920K3	Q920k3 rattus norv
6	738.5	24.6	492	4 Q96T73	Q96t73 homo sapien
7	712.5	23.8	767	13 Q9DGR2	Q9dgr2 xenopus lae
8	690.5	23.0	405	4 Q96E86	Q96e86 homo sapien
9	578	19.3	799	11 Q9DBI0	Q9dbi0 mus musculus
10	568.5	19.0	855	11 Q9JJI7	Q9jji7 rattus norv
11	543	18.1	643	6 Q97506	Q97506 sus scrofa
12	541.5	18.1	624	11 Q9DAT3	Q9dat3 mus musculus
13	536.5	17.9	624	11 Q91Y47	Q91y47 mus musculus
14	530.5	17.7	624	6 Q95ME7	Q95me7 oryctolagus
15	524	17.5	1379	5 Q9V4N6	Q9v4n6 drosophila
16	518.5	17.3	321	4 Q96R28	Q96r28 homo sapien

17	516.5	17.2	749	11 Q99JC8	Q99jc8 rattus norv
18	512	17.1	845	13 Q9DGR1	Q9dgr1 xenopus lae
19	508	16.9	812	11 Q9R0W3	Q9r0w3 rattus norv
20	502.5	16.8	339	11 Q99L44	Q99l44 mus musculus
21	499	16.6	812	11 Q91WJ5	Q91wj5 mus musculus
22	492.5	16.4	810	4 Q15146	Q15146 homo sapien
23	491	16.4	329	6 Q9GLI0	Q9gli0 ovis aries
24	488.5	16.3	273	11 Q921N4	Q921n4 mus musculus
25	487.5	16.3	581	5 Q9XZM7	Q9xmz7 stronglyloce
26	483	16.1	868	5 Q9Y1V3	Q9y1v3 polvandroca
27	479	16.0	415	6 Q29015	Q29015 sus sp. pre
28	477	15.9	273	6 Q9XSM1	Q9xsm1 ovis aries
29	476.5	15.9	334	6 Q46507	Q46507 papio hamad
30	475	15.8	275	4 Q96R26	Q96r26 homo sapien
31	474.5	15.8	787	5 Q9VEY6	Q9vey6 drosophila
32	471.5	15.7	317	13 Q9DGR3	Q9dgr3 xenopus lae
33	471.5	15.7	322	11 Q920S2	Q920s2 mus musculus
34	471.5	15.7	377	6 P79343	P79343 bos taurus
35	471	15.7	454	6 Q46506	Q46506 papio hamad
36	470.5	15.7	389	13 Q9FVX7	Q9fvx7 xenopus lae
37	469.5	15.7	282	11 Q9D4I3	Q9d4i3 mus musculus
38	468.5	15.6	1524	13 Q91674	Q91674 xenopus lae
39	466	15.5	421	11 Q60491	Q60491 cavia porce
40	460.5	15.4	297	11 Q88781	Q88781 rattus ratt
41	460.5	15.4	806	6 Q18783	Q18783 macropus eu
42	446.5	14.9	310	11 Q91XC4	Q91xc4 mus musculus
43	446	14.9	277	5 Q96899	Q96899 scolopendra
44	446	14.9	279	11 Q9QZ74	Q9qz74 rattus norv
45	441.5	14.7	310	11 Q9QY29	Q9qy29 mus musculus

ALIGNMENTS

RESULT 1
Q9BYE2 PRELIMINARY; PRT; 581 AA.
ID AC Q9BYE2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MEMBRANE-TYPE MOSAIC SERINE PROTEASE.
GN MSPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Shatmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC EMBL; AB048796; BAB39741.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.087; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001190; SRCR.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.


```
OS Homo sapiens (Human)..
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isoai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027798; BAB5376.1; -.
SQ SEQUENCE 491 AA; 52560 MW; 9D9D7FE308E593E6 CRC64;

Query Match 73.6%; Score 2207; DB 4; Length 491;
Best Local Similarity 96.2%; Pred. No. 1.5e-168;
Matches 426; Conservative 0; Mismatches 7; Indels 10; Gaps 2;

QY 1 MERDSHGNSPARTPSAGASPAQASPAQASPAQASPAQASPAQASPAQASPAQAS 60
DB 1 MERDSHGNSPARTPSAGASPAQASPAQASPAQASPAQASPAQASPAQASPAQAS 60
QY 61 PAGTTPGRASPGR-----ASPAQASPARASPALASLSRSSSGRSSASASVTTSPTRY 115
DB 61 PAGTTPGRASPGRASPAQASPAQASPARASPALASLSRSSSGRSSASASVTTSPTRY 120
QY 116 LVYRATPVGAVPIRSSPARSAPATRATRESPTSLPKFTWREGOKQLPGCVLLIALVV 175
DB 121 LVYRATPVGAVPIRSSPARSAPATRATRESPTSLPKFTWREGOKQLPGCVLLIALVV 180
QY 176 SLIILFQFQWQHTGIRYKQRESCPKHAYRCVDCGVDCCLKSDLGCVRFWDKSLKLIYS 235
DB 181 SLIILFQFQWQHTGIRYKQRESCPKHAYRCVDCGVDCCLKSDLGCVRFWDKSLKLIYS 240
QY 236 GSSHOWLPICSSNWNDSYSEKTCQQLGFSAHRTTEVAHRDFANSFSLRYNSTIQESLH 295
DB 241 GSSHOWLPICSSNWNDSYSEKTCQQLGFSAHRTTEVAHRDFANSFSLRYNSTIQESLH 300
QY 296 RSECPQSRYISLQCSHGLRAMTGRVVGALASDSKWPQVSLHFGTHICGGTLIDAOV 355
DB 301 RSECPQSRYISLQCSHGLRAMTGRVVGALASDSKWPQVSLHFGTHICGGTLIDAOV 360
QY 356 VLTAAHCFEFTREKVLGKVKYAGTSLNHLQPEAASIAEIIINSNTDEDDYDIALMRL 415
DB 361 VLTAAHCFEFTREKVLGKVKYAGTSLNHLQPEAASIAEIIINSNTDEDDYDIALMRL 420
QY 416 SKPLTLSAH-----IHPALCPMH 433
DB 421 SKPLTLSGREGICTPRSPAPQQR 443

RESULT 4
QY1VQ8 PRELIMINARY; PRT; 176 AA.
AC QY1VQ8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SIMILAR TO MOSAIC SERINE PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST TUMOR;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010843; AAH10843.1; -.
KW Protease.
```

```
FT NON_TER 1
SQ SEQUENCE 176 AA; 19809 MW; F3B8C573C07E9717 CRC64;

Query Match 30.8%; Score 925; DB 11; Length 176;
Best Local Similarity 94.3%; Pred. No. 1.8e-86;
Matches 166; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 387 PEASATIEIINSNYTDEDDYDIALMRLSKPLTSLSAHIHPACLPMHGQTFSLNETCWI 446
DB 1 PEASATISQIINSNYTDEDDYDIALRLSKPLTSLSAHIHPACLPMHGQTFGLNETCWI 60
QY 447 GFCKTRETDDKTSFPLREVQVNLIDFKKNDYLVYDYSYLTFRMWCAGDLRGDRSCQGS 506
DB 61 GFCKTRETDEKTSFPLREVQVNLIDFKKNDYLVYDYSYLTFRMWCAGDLRGDRSCQGS 120
QY 507 GGPLYVCEONNRWYLAGVTSWGTGCGQNKPGVYTKYTEVLPWLYSKMESEVRRPKS 562
DB 121 GGPLYVCEONNRWYLAGVTSWGTGCGQNKPGVYTKYTEVLPWLYSKMESEVRRPKS 176

RESULT 5
QY20K3 PRELIMINARY; PRT; 490 AA.
AC QY20K3;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TMRPS2.
GN TMRPS2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Tsuzuki S.;
RT "TMRPS2, Rat.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073550; BAB70683.1; -.
SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;

Query Match 25.5%; Score 763.5; DB 11; Length 490;
Best Local Similarity 33.0%; Pred. No. 5.6e-53;
Matches 182; Conservative 78; Mismatches 202; Indels 89; Gaps 16;

QY 27 AGTTPGRASPAQASPAQASPAQASPAQASPAQASPAQASPAQASPAQASPARASP 86
DB 5 SGSPPG-----IGPYENHGQSEHYVSP--RPPVSPSGYNLYPAQSCP----- 46
QY 87 ALASLSRSSSGRSSASASVTTSPTRVYLVRATPVGAVPIRSSPARSAPATRATRES 146
DB 47 -----SPVPQYAPRVTTQASTPAIHQPRSS--GTLCSTKS-- 80
QY 147 TSLPKFTWREGOKQLPGCVLLIALVSLI--ILFQFWQHTGIRYKQRES--CPKH 202
DB 81 -----KKSMLVALALGLSSPVAAGLLKWFWDKSCSSSEMGSGTCLISS 128
QY 203 AVRCDDGVDCCLKSDLGCVRFWDKSLKIYSGSSHOWLPICSSNWNDSYSEKTCQQLG 262
DB 129 SLWCDGVAQCPNGKDNRCVRLYGTFTLQVYSSQRKAMYPCVQDDWNESYGRAACKDMG 188
QY 263 FESAHTTE--VAHRDFANSFSLRY---NSTIQESLHRS--CPSQRYISLQCSHGLRAM 317
DB 189 YKNSFSSGQIPDQSGATSEFMKLNWAGNVLDYKLYHSDSCSSRWVSLRCLTEGVRV 248
QY 318 --TGRIVGALASDSKWPQVSLHFGTHICGGTLIDAOVLTAAHCFEFTREKVLG-- 373
DB 249 RQSRIVGGSTASPGDWPQVSLHVGQIHVGGSGIITPEWIVTARHCV-----EPLSSPR 304
QY 374 -WKVYAGT-----SNLHOLPEAASIAEIIINSNTDEDDYDIALMRLSLPLTLSAH 424
DB 374 -WKVYAGT-----SNLHOLPEAASIAEIIINSNTDEDDYDIALMRLSLPLTLSAH 424
```

```
Db 305 YWTAFAILKQSLMFYGSRRHQ-----VEKVISHPNYDSTKKNNDIALMKLQTLPAFNDV 358
QY 425 IHPACLPMHQGTFSLNETCWITGFKTRETDDKTSFPLREVQVNLIDFKKNDLYVDYSY 484
Db 359 VKPCLPNPGMLDLAQECWISGWATYE-KGKTSVDVLAAMVPLIEPSKNSKIYNNL 417
QY 485 LTPRMKAGDLRGDRSCQSGDGLVCEQNNRWYLAGVTSWGTGCGQRNKPVGYYTKVTE 544
Db 418 ITPAMICAGFLQSGVSDQSGDGLVTLKNEIWWLIGDTSWGSCKAKAYRPGVYGNVTV 477
QY 545 VLPWIYSKMS 555
Db 478 FTDWIYQOMRA 488

RESULT 6
Q96T73 PRELIMINARY; PRT; 492 AA.
AC Q96T73;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE EPIITHELIALIN.
GN THPRS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223025; PubMed=11322890;
RA Jacquelinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,
RA Hoidal J.R.;
RT "Cloning and characterization of the cDNA and gene for human
RT epitheliasin.";
RL Eur. J. Biochem. 268:2687-2699(2001).
DR EMBL; AF329454; AAK53559.1; -.
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DAF CRC64;

Query Match 24.6%; Score 738.5; DB 4; Length 492;
Best Local Similarity 31.9%; Pred. No. 5.7e-51;
Matches 178; Conservative 87; Mismatches 194; Indels 99; Gaps 18;

QY 27 AGTPPGRA-----SPAQASPAQASPAQTPGPRASPAQASPAQTPGPRASPAQ 78
Db 5 SGSPPAIGPYENHGYQENPYPAQTVVPT-VYEVHPAQYFSPVP-----QYAPRV 56
QY 79 ASPARASPALASLSSGSRSSARSASVTTSPTVYLVRATPVGAVPINSRSPASAPAT 138
Db 57 LTQA-SNPVCTQPKSPG-----TVCTSKTKALCITLTLGTF----- 94
QY 139 RATRESPTSLPFTWREGQKQLPLIGCVLLALVLSLILFQFWQ-----GHTGIRY 192
Db 95 -----LVGA-----ALAGLL-----WKFMGSKCSNGIEC 120
QY 193 KEORESCPKHAVRCGVVDCGLKSLGCVFDMWDLKLIYSGSHQMLPICSSNNDS 252
Db 121 -DSSGTCINPNCWCDGSHCPGEGDENRCVLYGPNFILQMYSSQRKSWHPVCQDDWEN 179
QY 253 YSEKTCQQLGFESAHTTEVAHRDFANSFILRYNST-----IQESLHRS-CPSORIYS 306
Db 180 YGRAACKDMGYKNKFYSQ-GIVDDSGSTSFMKLNTSAGNVIYKLYHSDACSSKAVYS 238
QY 307 LQCSHGCLR---AMTKRIVGGALASDKPWPQVSLHFGTTHICGGTLIDAQWVLTAAHCF 363
Db 239 LRCIAGVNLNSRQRIYGGESALPCAMPQVSLHVQNVHVCVCGSIITPEWIVTAHCV 298
QY 364 FVTRKVLG---WKYAGTASNLHQLPEAA--SIAPIIINSVNTDEDDYDIALMRLSKP 418
Db 299 -----ERPLNPNWHTAFAGILRQSFMYGAGYQVEKVISHPNPDYDSTKKNNDIALMKLQKP 354
QY 419 LTLSAIHHPACLPMHQGTFSLNETCWTGFKTRETDDKTSFPLREVQVNLIDFKKNDY 478
```

```
Db 355 LTNDLVKVPCLPNPGMLQPEQLCWTISGWGATEE-KGKTSEVLNAAKVLLIETQRNSR 413
QY 479 LVYDLSVLTPRMKGAGDLRGDRSCQSGDGLVCEQNNRWYLAGVTSWGTGCGQRNKPVG 538
Db 414 YVDNLITPAMICAGFLQGNVDSCQSGDGLVTSKNNIWWLIGDTSWGSCKAKAYRPGV 473
QY 539 YTKVTEVLPWIYSKMS 556
Db 474 YGNVMVFTDWIYQOMRAD 491

RESULT 7
Q9DGR2 PRELIMINARY; PRT; 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE EMBRYONIC SERINE PROTEASE-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
RT from Xenopus laevis.";
RL Gene 252:209-216(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00766; ICHG.
DR MEROPS; S01.049; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001190; SRCR.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00057; ldl_recept_a; 6.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 8.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; tryp_spc; 1.
DR PROSITE; PS50068; LDLRA_2; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 23.8%; Score 712.5; DB 13; Length 767;
Best Local Similarity 35.7%; Pred. No. 1.2e-48;
Matches 162; Conservative 63; Mismatches 176; Indels 53; Gaps 12;

QY 134 SAPATRATRESPTSLPKFTWREGQKQLPLIGCVLLIALLVLSLILFQFWQ----- 186
Db 333 TATTTTATTTTTPC-----QIYCMNFM--YYTCTIAYQCMCGVRQCYG 376
QY 187 -----HTGIRYKQR-----ESCPKHAVRCGVVDCGLKSLGCVFDMWDLKLIY 234
Db 377 DDELNCDTTITTAICRKCSSVSVLSQWCDGVSDCPYGEDEMSCVSLYPADFQLOVY 436
QY 235 SGSSHQWLPICSSNNDSYSEKTCQQLGFE-SAHRTTEVAHRDFA-NSFSIL-----R 285
Db 437 STSVSAWLPVCSYDWNDDGFRACQDFYNGSSYNNRYDTLMSYPADPNGYFKLYSGYWRSK 496
QY 286 YNSTIQESLHRSRSCPSQRIYSLQCSHGL--RAMTGRIVGGALASDKPWPQVSLHFGTT 343
```



```
QY 245 CSS-----NWDSYSEKTCOOLGFESAHRTTEVAHRDFANFSILRYNSTIQESLHRSEC 299
Db 498 CDROPDLNGSD---EQCOE-GVPCGTFTFOCEDR-----SCVKKP--NPEC 539
QY 300 PSORYI----SLQCSHGLRAMTGRIVGGALASDSKWPQVSLHFGTTHTCGFTLIDAOW 355
Db 540 DGSDCRDGDSDHCDGGLGSLRIVGGTVSGEGWPQASLIQGRHICGGLAIADRW 599
QY 356 VLTAACHFFVTRKVLGKWKVYAGTSLNLH-QLPEAAS--IAEIIINSNTYDEDDYDIAL 412
Db 600 VITAACHFCQEDSNASPKLWTVFLGKMRQNSRWPGEVSKVSRFLFLPHYHEEDSHDYDAL 659
QY 413 MRLSKPLTJLSAHTHACPLPMHGOTFSLNCTWITGKTKRETDDKTSPLFUREVOVNLIDF 472
Db 660 LQDHPVVSATVRPVCPLPARSHFFEPGQHCWITGWAQRE-GGPVSNLTQKVDVQLVPQ 718
QY 473 KKNNDVLVDSYLTPRMCAAGDLRGDRSCOGDGGPLVC-EQNNRWYLAGVTSWGTGG 531
Db 719 DLCE--AVRYQVPRMLCAGYRKGRKADACQDGGGGLVCRPSGRWFLAGLVSWGLGCG 776
QY 532 QRNKPVGYYTKVTEVLPWI 549
Db 777 RPNFEGYTRVTRVINWI 794

RESULT 10
Q9JJ17 PRELIMINARY; PRT; 855 AA.
AC Q9JJ17
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MEMBRANE BOUND SERINE PROTEASE (MEMBRANE BOUND ARGININE SPECIFIC
DE SERINE PROTEASE).
GN MBSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA TSUZUKI S.; TISSUB-JEJUNUM;
RC STRAIN=WISTAR; TISSUB=JEJUNUM;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUB=DUODENUM;
RA Inoue H., Takahashi K., Kishi K.;
RT "membrane-bound arginine specific serine protease.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR EMBL; AB037898; BAB03502.1; -;
DR EMBL; AB049189; BAB13765.1; -;
DR HSSP; P00763; IDPO.
DR MEROPS; S01.302; -.
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 2.
```

```
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
FT VARIANT 665 665 K -> N.
SQ SEQUENCE 855 AA; 94955 MW; 35806B7ECF6CF03D CRC64;

Query Match 19.0%; Score 568.5; DB 11; Length 855;
Best Local Similarity 34.5%; Pred. No. 4.7e-37;
Matches 141; Conservative 54; Mismatches 133; Indels 81; Gaps 17;

QY 199 CPKHAVRCQGVDPCKLKSDLGCVRFEDWCKSLKLIYSGSSHQWL-----P----ICSS- 247
Db 464 CIRKDLRCDGWADCPDYSERHC-----RCNATHQFMCKNQCKPLFWVCDV 511
QY 248 -NWDSYSEKTC-----QQLGFESAHRTTEVAHRDFANFSILRYN 287
Db 512 NDCGDSDEEGSCSPAGSEFKCSNGKCLPOSQOCNGKDDCGDGDSEASCDNVNAVSCTKYT 571
QY 288 STIQESL---HRSECPORYISLQCS-----HCGLRAMT--GRIVGGALASDSKWP 333
Db 572 YRCQNGLCNLKNGPCDGGK-----DCSDGSEKNCDCGLRSFTKQARVVGTTNADGEWP 627
QY 334 WQVSLH-FGTTHICGGLIDAQWLVTAACHF-----FVTRKYLEGMKVYAGTSNLHQLP 387
Db 628 WQVSLHALGQGHLCASLIISPDWLVSAAHCFQDETIFKYSHTM--WTAFLGL--LDQSK 683
QY 388 EAAS-----IABIINSNTYDEDDYDIALMRLSKPLTJLSAHTHACPLPMHGOTFSLNE 441
Db 684 RSASGVQEHKLRITHTPSFNDFTFDYDIALELEKPAEYSTVVRPCLPDNTHVFPAGK 743
QY 442 TCWITGFGKTRTDDKTSPLREVQVNLIDFKKCNLYVYDSYLTPRMCAAGDLRGDRS 501
Db 744 AIWVTGWGHTKE--GGTGALLIQKEIRVINQTTCELL--PQGITPRMCMCVGLSGVDS 800
QY 502 CQDGGSGPL-VCQNNRWYLAGVTSWGTGCGRNKPGVYTKVTEVLPWI 549
Db 801 CQDGGSGPLSSVEKGRIFQAGVSVSGEGCAQRNKPQVYTRIPEDRWI 849

RESULT 11
ID O97506 PRELIMINARY; PRT; 643 AA.
AC O97506;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA;
RA Kimura A., Kihara T., Okimura H., Hamabata T., Ohnishi J.,
RA Moriyama A., Takahashi K., Takahashi T.;
RT "Identification of porcine follipain as plasma kallikrein, and its
RT possible involvement in the production of bradykinin within the
RT follicles of porcine ovaries.";
RL Mol. Reprod. Dev. 57:79-87(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB022425; BAA37147.1; -.
DR HSSP; P00766; ICHG.
DR MEROPS; S01.212; -.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
```

```
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLIEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; TRYP-SPC; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS00240; TRYP-SPC; 1.
DR PROSITE; PS00134; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 643 AA; 72227 MW; AFF2923E3C3CB80A CRC64;

Query Match      18.1%; Score 543; DB 6; Length 643;
Best Local Similarity 32.1%; Pred. No. 3.6e-35;
Matches 130; Conservative 58; Mismatches 125; Indels 92; Gaps 13;

QY 212 CKLSDELGCVRFDMDKSLIKIYSGSHQWLPTICSSNNWDSYSEKTCQOLGFESAHT-- 269
DB 264 CFLKTSHTGTPSF-----PTQENALSGYSLLTCKQTLPEPCHSKIY 305

QY 270 TEVAHRDFANFSILRYNSTIOE-----SLRSECPQRYISLQCS----- 310
DB 306 SEVDFEGEELNFTVQGANLCQETCTKTIRCOFFTYSLHPEDCRGEK---CKCSLRSSD 362

QY 311 -----HGLRAMTG-----RIVGGALASDSKWPQVSLHF--- 340
DB 363 GSPTKITH-GMRASSGYSRLCRSDGHSACATYKANTRIVGGTDSFLGEPWPQVSLQAKLR 421

QY 341 GTTHICGGTLDIAQWVLTAAHCFVFTREKVLGE-----WKVYAGTNSLHQLPEA---AS 391
DB 422 AQNLHCGGSIIGHQWVLTAAHCF-----DGLSLPDILWRIYGGILNISEITKETPFESQ 473

QY 392 IAEIINSNYTDEDDYDIALMRLSKPLTSAHIHPACLPMMHGQTFSLNETCWITGFGKT 451
DB 474 VKEIITHQYKTLSEGHDIALLKLETLPLNYTFQKPICLPSRDDTNVVTYNCWVTGWTGFT 533

QY 452 RETDDKTSPLREVQVNLDFDKNDYLVDSYLPTRMMCAGLDRGRDSCOGDGGPLV 511
DB 534 EKEGE-IQNLQKVNIPVSNBECQK-SYRDHKISKQMICAGYKEGKDGACKGESGGLV 591

QY 512 CQONNRWYLAGVTSWGTGCGQGNKPGVYTKVTEVLPWYISKMESE 556
DB 592 CKYNGIHLVGTGTSWGECCARREQPGVYTKVTEYMDWILEKTQDD 636

RESULT 12
Q9DAT3      PRELIMINARY;      PRT; 624 AA.
AC Q9DAT3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 1600027G01RIK PROTEIN.
GN 1600027G01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PLACENTA;
RX MEDLINE=21085660; PubMed=41217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Queckenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

```
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AK005546; BAB24114.1; -.
DR HSSP; P00750; 1RTF.
DR MGD; MGI:1919281; 1600027G01RIK.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLIEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; TRYP-SPC; 1.
DR PROSITE; PS00240; TRYP-SPC; 1.
DR PROSITE; PS00134; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 624 AA; 69788 MW; OEDDEBC56009E97 CRC64;

Query Match      18.1%; Score 541.5; DB 11; Length 624;
Best Local Similarity 29.7%; Pred. No. 4.5e-35;
Matches 138; Conservative 66; Mismatches 150; Indels 111; Gaps 17;

QY 193 KEQRESCPKHVRCDGVDP-CKLKS---DELGCVREDWPKSLKLIYSGS----- 238
DB 168 KYTRTCTPTTITKLNGVWSGFSKGLNLACIRDFNPNTVLADNLNIDSVAPDAFVCR 227

QY 239 -----HOWLPIK-----SSNWDYSSEKTCQOLGFESAHTTEVAHRDFANFSILR--- 285
DB 228 RICTHH--PTCLFFTFFSQAMPKESQRHILCLKLTSEGLPSTRITKSHALSGFSLQCHR 285

QY 286 -----YNST--IQESL-----HRSEC-----PSQR----- 303
DB 286 SVPVFCPSFYNDTDFLGEELDIVVKGQETCQKTCCTNNARCOFFTYFSLRLCNERNR 345

QY 304 ---YISLQCSHCLRAMTG-----RIVGGALASDSKWPQV 336
DB 346 GRCYLKLSSNGSPTRILHGRGGISYSLRCKMDNVCTTKINPRVVGGAASVHGEPWQV 405

QY 337 SLHFGTHICGGTLDAQWVLTAAHCF--FVTRKVLGKWKYVAGTNSLHOLPEAS--- 391
DB 406 TLHISQHLGCGSIIGNQWILTAAHCFSGIETPKL----RVYGGIVNQSEINEGTAFFR 461

QY 392 IAEIINSNYTDEDDYDIALMRLSKPLTSAHIHPACLPMMHGQTFSLNETCWITGFGKT 451
DB 462 VQEMIHTDQYTTAESGYDIALKLESAMNYTFQKPICLPSKGDNRNAVHTECWVTGWTG 521

QY 452 RETDDKTSPLREVQVNLDFDKNDYLVDSY-LTPRMCAGLDRGRDSCOGDGGPL 510
DB 522 ALRGEVQST-LQAKVPLVSNBECQTR--YRRHKITNKMICAGYKEGKDGKTCGDGGPL 578

QY 511 VCEQNNRWYLAGVTSWGTGCGQGNKPGVYTKVTEVLPWYISKMES 555
DB 579 SKYNGVHLVGTGTSWGECCGQKREGVYTNVAKYVDWILEKTQT 623

RESULT 13
Q91Y47      PRELIMINARY;      PRT; 624 AA.
ID Q91Y47
```

```
AC Q91Y47;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COAGULATION FACTOR XI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.WR; TISSUE=LIVER;
RX MEDLINE=97385041; PubMed=9242536;
RA Gailani D., Sun M.F., Sun Y.;
RT "A comparison of murine and human factor XI.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF356627; AAK40233.1;
SQ SEQUENCE 624 AA; 69874 MW; 49D281BFAEC12A03 CRC64;

Query Match 17.9%; Score 536.5; DB 11; Length 624;
Best Local Similarity 29.7%; Pred. No. 1.1e-34;
Matches 138; Conservative 65; Mismatches 151; Indels 111; Gaps 17;

QY 193 KEQRESCPKHVRCDGVD--CKLKS---DELGCVRFWDKSLKLIYSGSS----- 238
Db KYTRTGTPTITKLNGVSGFSKSLGSLNLACIRIFNPTVLADNLIDSVAPDAFVCR 227
QY 239 -----HOWLPIC-----SSNWDYSSEKTCQQLGFESAHRTTEVAHRDFANSFSILR--- 285
Db 228 RICTHH--PCFLFFTFEQAWPKESQRHLCLLTKTSEGLPSTRIKHALSGSLQCHRH 285
QY 286 -----YNST--IQESL-----HRSEC-----PSQR----- 303
Db 286 SVPVFCHPSFYNDTFLGEELDIVDKGQETCQCTNARCOFTFYPSHRLCERNRR 345
QY 304 ---YISLQCHSGCLRAMTG-----RIVGALASDSKWPQV 336
Db 346 GRCLYLSSNGSPTRIHNGRGSGSLRLCKKMDNVCTKINRVVGGAAVHGEWPMQV 405
QY 337 SLHFGTHTCGGTFLDAQWVLTAAHCF--FVTRKVLGKWKVYAGTSLNHLQPEAAAI-- 392
Db 406 TLHISQHLGGGSIIGNQWILTAAHCFSGTETPKL---RVYGGIVNQSEINEGTAFPR 461
QY 393 -AEIINSNTDEDDYDIALMLRSLPLTLSAHIHPACLPMHGQTSLNCTCWITGFGKT 451
Db 462 EQEMIIHQVTTAESGYDIALLESAMNYTDFQRPICLPKSGDRNAVHTECVTGWGTY 521
QY 452 RETDDKTSPLREVQVNLIDFKKNDYLVYDSY-LTPRMACAGDLRGDCOGDSGGPL 510
Db 522 ALRGEVQST-LQAKVPLVSNESQTR--YRRKIINTKMKICAGYKEGGKDTCKDGGPL 578
QY 511 VCDONNRWYLAGVTSWGTGCGQRNKPQVYTKVTEVLPWYISKMES 555
Db 579 SKYNGVHLVGITSNGEGGQKRPQVYTNVAKYVDWILEKTQT 623

RESULT 14
Q95ME7 PRELIMINARY; PRT; 624 AA.
AC Q95ME7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DE 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE FACTOR XI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```

```
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Sinha D., Marcinkiewicz M., Gailani D., Walsh P.N.;
RT "Rabbit Factor XI and Homodimerization: Complete cDNA sequence without
a cDNA library.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF395821; AAK82432.1;
SQ SEQUENCE 624 AA; 69961 MW; A7A8851219386B6B CRC64;

Query Match 17.7%; Score 530.5; DB 6; Length 624;
Best Local Similarity 28.7%; Pred. No. 3.4e-34;
Matches 138; Conservative 63; Mismatches 157; Indels 123; Gaps 16;

QY 178 IILFQFMQ--GHTGIRYKEQ-----RESCPKHVRG-----DGWVCKLKLSDE 218
Db 163 ICULKYTQTCAPTGIKRLKAVSGFSKSLKSCALNLACIRIFFSTVFADNNIDSVAPDA 222
QY 219 LGCVRFDWKSLLKIYSGSSHOWLPIC-----SSNWDYSSEKTCQQLGFESAHRTTEV 272
Db 223 LVCRRY-----CTHH--PNCFLFFTFEQWPKESHRLNCLLTKTSEGLPSTRI 268
QY 273 AHRDFANSFSILRYNSTIQESLHRS-----ECPS 301
Db 269 HKQALSGFSLQNCRHSPVCHSSFYDFTFLGEELDIVDKGHEACQKMCSTAIRCQF 328
QY 302 QRYISLQCHS-----CGLRAMTG-----RIV 322
Db 329 FTVSSQESHNKKGTCYLKSSNGSPTKILHGRGGISGYTLRLCKMDNVCTTKIRPRIV 388
QY 323 GGALASDSKWPQVSLHFGT---THICGGTFLDAQWVLTAAHCF--FVTRKVLGKWKVYA 378
Db 389 GGSASLPGEWPMQVTLHTVSPTRHLCGGSIIGNQWILTAAHCFYGTSPKIL---RVYG 445
QY 379 GTSNLHLQPEAAS---IAEIIINSNTDEDDYDIALMLRSLKPLTLSAHIHPACLPMHGQ 435
Db 446 GILNQSEIKEDTAFVGQEIILIHQYKTAESGYDIALLETMTNYTDSORPICLSKGD 505
QY 436 TFSLNTCTWITGFGKTRTDDKTSPLREVQVNLIDFKKNDYLVYDSY-LTPRMACAGD 494
Db 506 RNVIYDCWVTGWG-YRKLDRDKIQLTQAKIPLLSNEEQKR--YORHEITSGMICAGY 562
QY 495 LRGRDCOGDSGGPLVCEONNRWYLAGVTSWGTGCGQRNKPQVYTKVTEVLPWYISKME 554
Db 563 KEGGKDACGDSGGPLSCRHNEVHLVGITSWEGCAQRPQGIYTNVYKYLWILEKTQ 622
QY 555 S 555
Db 623 A 623

RESULT 15
Q9V4N6 PRELIMINARY; PRT; 1379 AA.
AC Q9V4N6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG2105 PROTEIN.
GN CG2105.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
```


RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster K., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Hellman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPsin FAMILY.
 DR EMBL; AE003840; AAF59230.1; -.
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.0PA; -.
 DR FlyBase; FBgn0033192; CG2105.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR000024; Fz_domain.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; Trypsin.
 DR InterPro; IPR000217; Tubulin.
 DR Pfam; PF01392; Fz; 1.
 DR Pfam; PF00057; ldl_recept_a; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00063; FRI; 1.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp-SPC; 1.
 DR PROSITE; PS50038; Fz; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS50287; SRCR_2; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00227; TUBULIN; 1.
 KW Glycoprotein; Hydrolase; Serine protease.
 SQ SEQUENCE 1379 AA; 149489 MW; A593A9CC2167E4B7 CRC64;

Query Match 17.5%; Score 524; DB 5; Length 1379;
 Best Local Similarity 31.6%; Pred. No. 3,le-33;
 Matches 127; Conservative 66; Mismatches 155; Indels 54; Gaps 17;

QY 201 KHAVRCGVDCKLSDELGCVRF-----DWDKSLKITYSGSSHOWLPICSSNWINDSYS- 254

Db 943 KHI--CDGIIDCPYGDERNCLRLSERNGDVGTVLEVYRIGORQMPACVKNDRAVSP 1000
 QY 255 EKTCCQQLGFESAHT---TEVAHRDFANFSI-----LRYNSTIQESLRHSECPQS 302
 Db 1001 SAVCSILGYSAVNATSVLTQLTHRPLLATVNVSTDIWKMYAKRKSTLMQEFANCKKTEDY 1060
 QY 303 RYISLQCSH--CGLRAMTG-----RIVGGALASDKWPQVSLHFGTTHI--CGGTLD 352
 Db 1061 PMADLTCSNYECG-RYKRGHFKPSRIIGTQASPGNWFFLAAILGGPEKIFYCAGVLIS 1119
 QY 353 AOMVLFAAHCFVYTRKVLKGVKQVYAGTSNLHOLP---EASTAEIIINSNYTDE-EDDY 408
 Db 1120 DQWVLTASHCVGNYSVIDLEDWTIQLGVTFRNSFTYSGQKVKVKAIPHPOYNNALAHDN 1179
 QY 409 DIALMELSKPLTLSAIIHPACL------MHGOTFSLNETCWITGFGKTRTDKTS--P 460
 Db 1180 DIALFOLATRVAFHEHLPLVCLPPPSVRNLHPGTL-----CTVIGWKREDKDKPKSTY 1234
 QY 461 FLREVQVNLIDFKKNDYLVYDSY-LTPRMWCAGDLRGGRDSCQDGGGGLVCE--QNN 516
 Db 1235 IYNEVQVPIITRNQCDLW--DNLTVSEGNVCAGFDGDKDACQDGGGLLCYPGKRN 1292
 QY 517 RWYLAGVTSWGTGCGQGRNKPQVYTKYTEVLPWITYSKMESEVR 558
 Db 1293 RWFVGIVSWGIMCAHPRLPGVYVNVVYVPMIQEQIAKHSR 1334

Search completed: August 20, 2002, 04:22:06
 Job time: 2382 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 16:02:42 ; Search time 3205.81 Seconds
(without alignments)
11410.404 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctccagaccatgagagagg.....ggctgctgtgactcgagaaa 1748

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

ALIGNMENTS

RESULT	1	AXI49579	Sequence 3 from Patent WO0136604.	1689 bp	DNA	linear	PAT 08-JUN-2001
LOCUS	AXI49579	Sequence 3 from Patent WO0136604.					
DEFINITION	AXI49579	Sequence 3 from Patent WO0136604.					
ACCESSION	AXI49579	Sequence 3 from Patent WO0136604.					
VERSION	AXI49579.1	GI:14348013					
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	1 (bases 1 to 1689)						
AUTHORS	Madison,E.L. and Ong,E.O.						
TITLE	Nucleic acids encoding endothelialases, endothelialases and uses thereof						
JOURNAL	Patent: WO 0136604-A 3 25-MAY-2001;						
FEATURES	CORVAS INTERNATIONAL, INC. (US)						
source	Location/Qualifiers						
	1. .1689						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	1. .1689						
	/note="DNA encoding a transmembrane serine protease"						


```

KEYWORDS      Homo sapiens lung cDNA to mRNA, clone_lib:Marathon Ready cDNA
SOURCE        (Clontech).
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     Kim,D.R., Sharmin,S., Inoue,M. and Kido,H.
AUTHORS       Cloning and expression of novel mosaic serine proteases with and
TITLE         without a transmembrane domain from human lung(1)
JOURNAL       Biochim. Biophys. Acta 1518 (1-2), 204-209 (2001)
MEDLINE       21167393
REFERENCE     2 (bases 1 to 2393)
AUTHORS       Kim,D.R., Inoue,M. and Kido,H.
TITLE         Direct Submission
JOURNAL       Submitted (12-SEP-2000) Hiroshi Kido, The University of Tokushima,
              Division of Enzyme Chemistry, Institute for Enzyme Research,
              Kuramoto-cho, Tokushima 770, Japan
              (E-mail:kido@ier.tokushima-u.ac.jp, Tel:81-88-633-7424,
              Fax:81-88-633-7425)
FEATURES      Location/Qualifiers
              1..2393
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /tissue_type="lung"
               /clone_lib="Marathon Ready cDNA (Clontech)"
              1..87
               /gene="mspl"
               /db_xref="GI:13429970"
               /transl_except=(pos:1573..1575,aa:His)
               /transl_except=(pos:862..864,aa:Arg)
               /transl_except=(pos:979..981,aa:His)
               /product="membrane-type mosaic serine protease"
               /protein_id="BAB39741.1"
               /db_xref="GI:13429970"
               /translation="MERDSHGNSAPARTPSAGAPAPASPAAGTPGRASPAQASPAQ
              SPAGTPGRASPAQASPAAGTPGRASPCGRASPAQASPARASPALASLSRSSGRSSSA
              RSATVPTSPVLRVLRATPGVAPLRSSPARSATRATRESPTSLPKFTWREGQKQ
              LPLGCVLLIALVVLILILQFQWGHGIRHKQRESCPKHVRCDGVDCUKLSD
              LGVRFWDKSLKLIYSGSHOWLPICSSNNDSYSEKTRQLGFESAHRTVEVAHRD
              FANFSILRYNSTQIESLHRSHCPQRYISLQCSHGLRAMTGRIVGALASDSKWPW
              QVSLHFHTICGGTLDIAQWLVAAHCFFVTRKLEGVKVIYAGTSLNHLQPPAASI
              AEIITNSNTDEEDYDIALMRLSKPLTSAHIHPACLPMHGOTFSLNETCWTITGFGK
              TREYDDKTSPLREVOVNLIDFKKNDYLVYDSVLTFRMCAGDLHGGRDSCQGDSSG
              PLVEQNNRWYLAGVTSWGTGCGQRNRPVYTKVTEVLPWLYISMEVRSLQQDTAPSR
              LGTSSGDPGGAPRV"
              1834..2393
              polyA_signal
              BASE COUNT 553 a 692 c 627 g 521 t
              ORIGIN
              Query Match      95.6%; Score 1670.4; DB 9; Length 2393;
              Best Local Similarity 99.9%; Pred. No. 0;
              Matches 1671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
              QY 1 ctcagagaccatggagaggagacagccaggggaatgcatactccagcaagaacaccttcagc 60
              DB 78 CTCAGAGACCATGGAGAGGAGACAGCCAGGGAATGCATCTCCAGCAAGAACACCTTCAGC 137
              QY 61 ttgagcatctcagccagggcatctccagctggagacacctccagccgggcatctccagc 120
              DB 138 TGGAGCATCTCAGCCCCAGGATCTCCAGCTGGGACACCTCCAGGCCGGGCATCTCCACG 197
              QY 121 ccaggcatctcagccagggcatctccagctggagacacctccggccgggcatctccagc 180
              DB 198 CCAGGCATCTCAGCCCCAGGATCTCCAGCTGGGACACCTCCGGCGGGGCATCTCCAGC 257
              QY 181 ccaggcatctcagctgttacacctccagggcggaatctccagctccagccgggcatctccagc 240
              DB 1338 GCCCCTGACCCCTGCGCTCAGATCCACCTGCTTGCCTTCCCTCCCTCCCTGATGATGACACCTTT 1397

```

QY	1321	tagcctcaatgagacctgtggtacacaggttttggcaagaccagggagacagatgacaa	1380
Db	1398	TAGCCTCAATGAGACCTGCTGGATCAGAGCTTTGGCAGAGACCAGGAGACATGACAA	1457
QY	1381	gacatcccccttctccgagagtgaggtcactcatctcatctcagctcaagaaatgcaatga	1440
Db	1458	GACATCCCCCTTCTCCGGAGGTGAGGTCAATCTCATCGACTTCAAGAAATGCAATGA	1517
QY	1441	ctacttgggtctatgacagttaccttaccccaagatgatgtgctgggacctctgtg	1500
Db	1518	CTACTTGGTCTATGACAGTTACCTTACCCCAAGATGATGTGCTGGGACCTTCTGTG	1577
QY	1501	ggcgagagactctcgccagggagacagcgggggtcttctgtgtgagcagaaacog	1560
Db	1578	GGGCAGAGACTCTCGCCAGGAGACAGCGGGGGCCTTGTCTGTGAGCAGAAACCG	1637
QY	1561	ctggacctggcagggtgtcacacagctggggacagagctgtgcccagagaaaaacctgg	1620
Db	1638	CTGGTACCTGGCAGGTGTCAACAGCTGGGCGACAGAGCTGTGGCAGAGAAACAACTGG	1697
QY	1621	tgtgtacacaaagtacagaagttctccctggatttacacagaatgag	1672
Db	1698	TGTGTACACAAAGTGACAGAAGTTCTTCCCTGGATTACAGCAAGATGGAG	1749
RESULT	3		
LOCUS	AX149581	2067 bp	DNA linear PAT 08-JUN-2001
DEFINITION	Sequence 5 from Patent WO0136604.		
ACCESSION	AX149581		
VERSION	AX149581.1	GI:14348015	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 2067)		
JOURNAL	Madison, E.L. and Ong, E.O.		
FEATURES	Nucleic acids encoding endotheliases, endotheliases and uses thereof		
source	Patent: WO 0136604-A 5 25-MAY-2001;		
CDS	CORVAS INTERNATIONAL, INC. (US)		
	1..2067		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	1..2067		
	/note="DNA sequence encoding a transmembrane serine protease (endotheliasase-L) protein"		
	/codon_start=1		
	/protein_id="CAC41221.1"		
	/db_xref="GI:14348016"		
	/translation="MERDSHGNSPARTSPASAPQASPAQSPACTPPGRASPAQASPAQA		
	SPAGTPPGRASPAQSPAGTPPGRASPAQASPARASPALASLSRSSGSSSA		
	RSAGVTTPTRVLYRATPVGAVPIRSSPARSPATRATRESPTSLPKFTWREGQKQ		
	LPLICGLVALLVSLILFLQFWQHTGIRYKEQRECPKHAVRCDGVDCKLKSD		
	LCQVRDMDKLLKITYSGSHOWLPICSNNWDSYSEKTCOOLGPESAHRTTEVAHRD		
	FANSPSILRYNSTIOESLHRSECPSORVISLQCSHGLRAMTGRIVGGALASDSKPWP		
	QVSLHFGTHICGGFLIDAQVLTAAHCFVYTRKVLGKVVYAGTSLNHLQPEASII		
	ABIIINSYNTDEEDYDIALMLRSLPLSLSAIHIPACLPMHGQTFSLNETCWTGFGK		
	TRETDDKTSFPLREVQVNLIDFKKNDYLVDSYLTPRMCCAGDLRGDRDSCQSGGG		
	PLVQCNRNRYLAVGTSWGTGQRNKQGVYTKTEVLPFIY SKMENRAQRVEKAWY		
	RPGRQLLGRCSPRSIFLCKVAMDFENVSIEDFVIFVKILCMGIRSSWPFPAFV		
	LVFLIFLILLLSFLUNKTSDSLTLTFTTAVTMRPLPENYHSPFPPLKIFHQLNLTIIK		
	EKKVIK"		
BASE COUNT	463 a 604 c 540 g 460 t		
ORIGIN			
Query Match	95.2%;	Score 1663.4;	DB 6; Length 2067;
Best Local Similarity	99.9%;	Pred No. 0;	
Matches 1664;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;

Db	301	CTTCCATCTCTAGCAGCAACTGGAATGACTCTCTACTCAGAGAAGACCTGCCAGCAGCTG	360
Qy	794	ggtttcagagagtgctcaccgacacacgaggttgccacagggatttgccacagcttc	853
Db	361	GGTTTCGAGAGTGTCTACCGACACACCGAGGTGCCACACAGGATTTTGCCACACAGCTTC	420
Qy	854	tcaatttgagatacaactccaccatccaggaagcctccacaggtctgaatgccttcc	913
Db	421	TCAATCTTGAATACAACTCCACCATCCAGAAAGCCCTCCACAGTCTGAATGCCCTTCC	480
Qy	914	cagcggatatactcctccagtgctccactgcgagactgagggccatgaccgagcgaac	973
Db	481	CAGCGGTATATCTCCCTCAGTGTCCCACTGCCGATGAGGGCATGACCGGGCGATC	540
Qy	974	gtggaggggctggcctcgatagcaagtggccttggcaagtggatgctgcacttcggc	1033
Db	541	GTGGAGGGGGCTGGCTCGGATAGCAAGTGGCTTTGGCAAGTGAGTCTSCACTTCGGC	600
Qy	1034	accaccacatctgtgagggcagctcattgacgcccagtggtgtcactgcgcgcccac	1093
Db	601	ACCACCCACATCTGTGAGGCACGCTCATTTGACGCCAGTGGGTCTCACTTGGCGGCCAC	660
Qy	1094	tgcttcttgtagccgggaaaggtcctgagggctggaaggtgtacgcgggcaccagc	1153
Db	661	TGCTTCTGTGTGACCCGGGAGAGGTCTTGAGGGCTGGAAGGTGTACCGGGGCCACGAG	720
Qy	1154	aacctgcaccagttgctgagcgagctccattgcccagatcatcaacagcaattac	1213
Db	721	AACCTGCACCACTTGCCTGAGCGAGCTCCATTGCCGAGATCATCAACAGCAATTAC	780
Qy	1214	accgatgagggagcagactatgacatgcgcccctcatgcggtgtgtccagcccctgacctg	1273
Db	781	ACCGATGAGGAGGACGACTATGACATGCCCTCATGCGGCTGTCTCAAGCCCTGACCTG	840
Qy	1274	tcgctcaatccaccctgttgcctcccattgcatgagacacatttagcctcaatgag	1333
Db	841	TCCGCTCACAATCCACCTGTGTGCTTCCCATGCCATGATGGACAGACTTTAGCGCTCAATGAG	900
Qy	1334	acctgtggatcaacaggtttggcaagaccagggagagacagatgacagacatcccccttc	1393
Db	901	ACCTGTGGATCACAGGCTTTGGCAAGACCAGGGAGACAGATGACAGACATCCCTCTC	960
Qy	1394	ctcgggaggtgcaggttcaatctcatcagacttcaagaaatgcaatgactacttggctat	1453
Db	961	CTCGGGAGGTGACAGTCAATCTCATCTCATCTCAAGAAATGCAATGACTACTTGGTCTAT	1020
Qy	1454	gacagttacattaccacaaagatgagtgtgtggtgggacattcgtggggcagagactcc	1513
Db	1021	GACAGTTACCTTACCCCAAGGATGATGTGTCTGGGGACCTTCTGTGGGGCAGAGACTCC	1080
Qy	1514	tgcagggagacagcgggggctctgtctgtgagcagaacacccgtgtgacttgga	1573
Db	1081	TGCCAGGAGACACGCGGGGGCTTGTGTGTGTGAGCAGAACACCCGCTGGTACTGGCA	1140
Qy	1574	ggtgtaccagctggggcacaggtgtgtggcagagaaacaaacctgtgtgtacacaaa	1633
Db	1141	GGTGTACCAAGTGGGACACAGGTGTGTGGCCAGAGAAACAAACCTGGTGTGTACACAAA	1200
Qy	1634	gtgacagaattcttccctggatttacagcaagatgagagcgaggtgcgattcagaaaa	1693
Db	1201	GTGACAGAAGTCTTCCCTGGATTTACAGCAAGATGAGAGCGAGGTGCGATTACAAAA	1260
Qy	1694	tcctaacagctggcctgtcgtctgcacagcaccgctgctgtga	1739
Db	1261	TCCTAACCAAGTGGCTTGTCTGTGCACAGCACCGGCTGCTGTGA	1306

RESULT 6
AX190546/c
LOCUS
DEFINITION
AX190546
ACCESSION
AX190546.1 GI:15143857

KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 1314)
AUTHORS	Quinn,K.E., Spytek,K.A., Majumder,K., Vernet,C., Burgess,C., Fernandes,E., Taupier,R.J., Rastelli,L. and Herrmann,J.L.
TITLE	Serine/threonine kinase and serine protease polypeptides and nucleic acids encoding same
JOURNAL	Patent: WO 0136645-A 17 25-MAY-2001;
FEATURES	Curagen Corporation (US)
source	Location/Qualifiers 1..1314
BASE COUNT	266 a 376 c 384 g 288 t
ORIGIN	
Query Match	74.5%; Score 1302.8; DB 6; Length 1314;
Best Local Similarity	99.8%; Pred. No. 3.3e-239;
Matches 1304; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Qy	434 agggagagccaggtacgagcctgcccaagttcacctggcgaggagggccagaagcagcta 493
Db	1314 ATGGAGAGCCAGGTACGAGCCTGCCAAGTTCACCTGGCGGAGGGCCAGAAGCAGCTA 1255
Qy	494 ccgctcaggggtgcgtgcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 553
Db	1254 CCGCTCATCGGGTGCCTCCTCATTTGCCCTGTGGTTCGCTCATCATCCTCTTC 1195
Qy	554 cagttctggcagggccacacagggatcaggtacaaagagcagagagagctgcccaag 613
Db	1194 CAGTTCGGCAGGGCCACACAGAGGATCAGGTACAGGAGCAGAGGAGAGCTGCCAAG 1135
Qy	614 cagcgtctgcgtgtgacgggggtggtgagctgcaagctgaagatgacagctgggctgc 673
Db	1134 CACGCTGTCTGCTGTACGGGGTGTGTGACTGCAAGCTGAAGAGTACGAGCTGGGTGC 1075
Qy	674 gtgaggttgactgggcaagctctcgtctaaatcactcctcctcctcctcctcctcctcctc 733
Db	1074 GTGAGGTTGACTGGGCAAGTCTCTGTTAAATCTACTCTGGGTCTCTCCATCAGTGG 1015
Qy	734 ctccccctctgtacagcaactggaatgactcctactcagagaagacgtgccagcagctg 793
Db	1014 CTTCCTCCATCTGTACACCACTGGAAATGACTCTCTACTCAGAGAAGAGCTGCCACAGCTG 955
Qy	794 ggttcgagagtgctcctcccgacacacgaggtgtgccacacagggattttgccaacagcttc 853
Db	954 GGTTTCGAGAGTGTCTACCGGACAACCGAGGTTGCCACAGGGATTTTGGCAACAGCTTC 895
Qy	854 tcaatcttgagatacaactccaccatccaggaagcctccacaggtctgaatgcctctcc 913
Db	894 TCAATCTTGAATACAACTCCACCATCCAGAAAGCCCTCCACAGGTCTGAATGCCCTTCC 835
Qy	914 cagcggatatactcctccagtgctccactgcgagactgagggccatgacggcgagcgc 973
Db	834 CAGCGGTATATCTCTCCCTCAGTGTTCCTCACTGCGGACTGAGGGCCATGACCGGGCGGATC 775
Qy	974 gtggaggggctgtgcctcggatagcaagtggccttggcaagtgtgactgtcgtcactgcgc 1033
Db	774 GTGGAGGGGGCTGTGGCTCGGATAGCAAGTGGCTTTGGCAAGTGTGACTGTGCATTCGGC 715
Qy	1034 accaccacatctgtgagggcagcgtcattgaagcccagttgggtgctcactgcgcgcccac 1093
Db	714 ACCACCCACATCTGTGAGGACGACGCTCATTTGACGCCAGTGGGTGCTCACTTGGCGGCCAC 655
Qy	1094 tgcttcttgtagccgggagaaaggtcctgagggctgggaaggtgtacgcgggcccagc 1153
Db	654 TGCTTCTGTGACCCGGGAGAGGTCCTGTGAGGGCTGTGAAGGTGTACGCGGGCACAGC 595
Qy	1154 aacctgcaccagttgcctgagggcagcctccattgcccagatcatcatcaacagcaattac 1213

[illegible]

Db 58 AAATCTTAACAGCTGGCTGCTGTCTGCACAGACACCGGCTGCTGTGA 9

RESULT 9
 LOCUS BC010843 2125 bp mRNA linear ROD 17-JUL-2001
 DEFINITION Mus musculus, similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
 ACCESSION BC010843
 VERSION BC010843.1 GI:14789982
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 2125)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgabps-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: http://www.shgc.stanford.edu
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAK Plate: 7 Row: b Column: 18
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source Location/Qualifiers
 1..2125
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3490022"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
 /clone_lib="NCI_CGAP_Mam5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 <1..531
 /codon_start=1
 /product="Similar to mosaic serine protease"
 /protein_id="AAH10843.1"
 /db_xref="GI:14789983"
 /translation="PEAASISQIILNGYITDEQDYDIALRLSKPLTSLNHIAPCL PMHGOTGLNETWITGFKETDEKTSPLREYQVNLDFKCNLYLIDSLTPR MMACGDLRGGRDSCGSGGPLVCEQNNRFLAGVTSWGTGCGGNKPGVYTKVTEVL PWYRKESEVFRFRKS"
 BASE COUNT 536 a 502 c 541 g 546 t
 ORIGIN

Query Match 25.5%; Score 446; DB 10; Length 2125;
 Best Local Similarity 89.7%; Pred. No. 2.1e-75;
 Matches 479; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1169 cctgaggcagctccatgccagatcatcatcaacagcaattacacgcagtaggagac 1228

Db 1 CCCGAGGCTGCTTATCTCCAGATCATCATCAACGCACTACACAGATGAACAGAT 60

Qy 1229 gactatgacatgcacctcagtgccgtgtccaaagccctgacccctgtccgtctcatccac 1288

Db 61 GACTATGACATTCGCCCTCATCAGGCTGTCCAAGCCCTGACCTGTGTCAGTTCATCATC 120
 Qy 1289 cctgtgctccctcccatgcatgacagacaccttttagctcctaatgagacctgtggtatcaca 1348
 Db 121 CCTGCTGCCTCCCGATGACGCGTCAGACCTTCGGCTCTCAATGAGACCTGTGGATCAG 180
 Qy 1349 ggtttggcaagaccagggagagacagatgacaagacatcccccttctccgggaggtcag 1408
 Db 181 GGCTTTGGCAAAACCAAGGAGACAGATGAGAAGACATCTCCCTTCCTCCGAGAGGTTT 240
 Qy 1409 gtcaatctcatcgacttcaagaatacaatgactacttggtctatgacagttaacctacc 1468
 Db 241 GTCAACTCATCTGACTTCAAGAAGTGCATGACTACTTGGTCTATGACAGCTACCTTACC 300
 Qy 1469 ccaagatgatgtgtgctggggacacctctggtggggcagagactccttgcacagggagacag 1528
 Db 301 CCAAGATGATGTGTCGGGGGATCTACGAGGAGGAGGAGGACTCTCCGACGAGGAGACGT 360
 Qy 1529 gggggccctctgtctgtgagcagaacacccgctgtacctgacaggtgtcaccagctgg 1588
 Db 361 GGAGGACCTCTCGTCTGTGAGCAGAAACAATCGCTGTGCTGCTGACGTGTCACGAGTGG 420
 Qy 1589 ggcacaggtgtggccagagaaacacacctggtgtgtatcaccaaaagtgcagaagtctt 1648
 Db 421 GGCACAGGCTGTGGCCAGAAACAAGCCTGTGTGTATACCAAAAGTGACAGAAAGTACTT 480
 Qy 1649 ccttgattcagcagaagtgtgagcagcaggtgtgctgattcagaaaatccctaaacca 1702
 Db 481 CCCTGGATTTACAGAAAGATGAGAGTGTGAGTGTGAGTGTGCGGAATCTTAACCA 534

RESULT 10
 AP002962/c 160222 bp DNA linear PRI 18-SEP-2001
 LOCUS AP002962
 DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-652L8, complete sequence.
 ACCESSION AP002962
 VERSION AP002962.2 GI:15637092
 KEYWORDS HTG.
 SOURCE Homo sapiens DNA, clone:RP11-652L8.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 160222)
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Homo sapiens genomic DNA
 JOURNAL Published Only in Database (2000) In press
 REFERENCE 2 (bases 1 to 160222)
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (28-NOV-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT On Sep 17, 2001 this sequence version replaced gi:11526589.
 FEATURES Location/Qualifiers
 1..160222
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-652L8"
 BASE COUNT 44819 a 37318 c 37151 g 40934 t
 ORIGIN

Query Match 22.6%; Score 394.8; DB 9; Length 160222;
 Best Local Similarity 96.1%; Pred. No. 1.1e-65;
 Matches 421; Conservative 0; Mismatches 2; Indels 15; Gaps 1;

QY 31 gaatgcattccagcaagaacaccttcagctggagcatctccagcccaggcatctccagc 90
|||||
Db 33836 GAATGCATCTCCAGCAAGAACCCTTCACTGGAGCATCTCCAGCCAGGCATCTCCAGC 33777
QY 91 tgggacacctccagccggcgcatctccagcccaggcatctccagcccaggcatctccagc 150
|||||
Db 33776 TGGGACACCTCCAGGCCGGGCATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGC 33717
QY 151 tgggacacctccagccggcgcatctccagcccaggcatctccagcccaggcatctccagc 210
|||||
Db 33716 TGGGACACCTCCAGGCCGGGCATCTCCAGCCAGGCATCTCCAGCTGGTACACCTCCAGC 33657
QY 211 ccgggcatctccagccg-----ggcatctccagcccaggcatctccagc 255
|||||
Db 33656 CCGGGCATCTCCAGGCCGGGCATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGC 33597
QY 256 ccgggcatctccggctctggcatctctccaggtctccagtcctccagtcctccagtcctccagc 315
|||||
Db 33596 CCGGGCATCTCCGGCTCTGGCATCACTTCCAGGTCTCTCCAGGTCTCTCCAGGTCTCTCCAGC 33537
QY 316 caggctcagcctcggtgacaacctcccaaccagagtgacctgttagagcaaacaccagt 375
|||||
Db 33536 CAGGTACGCTCGGTGACAACTCCCAACAGAGTGACCTTGTGTAGAGCAACACCACT 33477
QY 376 gggggctgtaccatccagcatctctcggcaggtcagcaccagcaaccagggccaccag 435
|||||
Db 33476 GGGGCTGTACCATCCGATCATCTCTCCAGGTGACAGGCATCCAGCCAGGCATCCAGC 33417
QY 436 ggagagcccaggtacag 453
|||||
Db 33416 GGAGAGCCCAGGTAAGG 33399

RESULT 11
AP000757/c
LOCUS AP000757 184926 bp DNA linear PRI 22-FEB-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-728F11,
complete sequences.
ACCESSION AP000757
VERSION AP000757.4 GI:13094221
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-728F11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Human genes and gene expression products
JOURNAL Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
REFERENCE 2 (bases 1 to 184926)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
COMMENT (E-mail:hattori@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Feb 21, 2001 this sequence version replaced gi:11094157.
FEATURES
Location/Qualifiers
source
1. 184926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-728F11"
BASE COUNT 47261 a 47596 c 43757 g 46312 t
ORIGIN
Query Match 22.6%; Score 394.8; DB 9; Length 184926;

Best Local Similarity 96.1%; Pred. No. 1.le-65;
Matches 421; Conservative 0; Mismatches 2; Indels 15; Gaps 1;
QY 31 gaatgcattccagcaagaacaccttcagctggagcatctccagcccaggcatctccagc 90
|||||
Db 163269 GAATGCATCTCCAGCAAGAACCCTTCACTGGAGCATCTCCAGCCAGGCATCTCCAGC 163210
QY 91 tgggacacctccagccggcgcatctccagcccaggcatctccagcccaggcatctccagc 150
|||||
Db 163209 TGGGACACCTCCAGGCCGGGCATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGC 163150
QY 151 tgggacacctccggcgccggcgcatctccagcccaggcatctccagctgtgtacacctccagc 210
|||||
Db 163149 TGGGACACCTCCGGCGCGGCATCTCCAGCCAGGCATCTCCAGCTGGTACACCTCCAGC 163090
QY 211 ccgggcatctccagccg-----ggcatctccagcccaggcatctccagc 255
|||||
Db 163089 CCGGGCATCTCCAGGCCGGGCATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGC 163030
QY 256 ccgggcatctccggctctggcatctctccaggtctccaggtctccaggtctccaggtctccagc 315
|||||
Db 163029 CCGGGCATCTCCGGCTCTGGCATCACTTCCAGGTCTCTCCAGGTCTCTCCAGGTCTCTCCAGC 162970
QY 316 caggctcagcctcggtgacaacctcccaaccagagtgacctgttagagcaaacaccagt 375
|||||
Db 162969 CAGGTACGCTCGGTGACAACTCCCAACAGAGTGACCTTGTGTAGAGCAACACCACT 162910
QY 376 gggggctgtaccatccagcatctctcggcaggtcagcaccagcaaccagggccaccag 435
|||||
Db 162909 GGGGCTGTACCATCCGATCATCTCTCCAGGTGACAGGCATCCAGCCAGGCATCCAGC 162850
QY 436 ggagagcccaggtacag 453
|||||
Db 162849 GGAGAGCCCAGGTAAGG 162832
RESULT 12
AX246661
LOCUS AX246661 450 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 1591 from Patent WO0166753.
ACCESSION AX246661
VERSION AX246661.1 GI:15861335
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 450)
AUTHORS Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
Stache-Crain,B.
TITLE Human genes and gene expression products
JOURNAL Patent: WO 0166753-A 1591 13-SEP-2001;
Chiron Corporation (US); Hyseq Inc. (US)
FEATURES
Location/Qualifiers
source
1. 450
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 97 a 166 c 113 g 74 t
ORIGIN
Query Match 19.0%; Score 332.8; DB 6; Length 450;
Best Local Similarity 94.3%; Pred. No. 9.4e-54;
Matches 362; Conservative 0; Mismatches 7; Indels 15; Gaps 1;
QY 1 ctccagagaccatgagagggagcagccagcgaatgcatctccagcaagaacacctccagc 60
|||||
Db 65 CTCAGAGACCATGGAGAGGGACGCCACGGGAATGCATCTCCAGCAAGAACACCTTCAGC 124
QY 61 tgggacattccagcccaggcatctccagctgggacacctccagccgggcatctccagc 120

```
|||||
Db 125 TGGACATCTCCAGCCAGGATCTCCAGCTGGGACATCTCAGCCGGGATCTCCAGC 184
Qy 121 ccaggcatctccagccaggcatctccagctgggacacctccggccgggcatctccagc 180
Db 185 CCAGGCATCTCCAGCCAGGATCTCCAGCTGTGACACCTCCGGCCGGGATCTACAGC 244
Qy 181 ccaggcatctccagctgtacacctccaggccgggcatctccaggccg----- 228
Db 245 CCAGGCATCTACAGCTGTGTACACCTCCAGGCGGGCATCTCCAGCCGGGATCTCCAGC 304
Qy 229 ---ggcatctccagccaggcatctccagccgggcatctccggctctggcatcttc 285
Db 305 CCAGGCATCTCCAGCCAGGATCTCCAGCCGGGATCTCCAGCTGTGCGCTCTGGCATCTTC 364
Qy 286 caggtctctccggcagggtcatctccgacgctcagctcggtagacacctccccaac 345
Db 365 CAGTCTCTATCCGGCAGGATCATCTCCGCGAGTCAAGCTCGGTGTAACCTCCCAAC 424
Qy 346 cagagtgtacctgttagagcaac 369
Db 425 CAGAGTACTCTGTTAGTAAAC 448

RESULT 13
BC012752
LOCUS
DEFINITION
Homo sapiens, similar to transmembrane protease, serine 4, clone
IMAGE:3623466, mRNA, partial cds.
ACCESSION
BC012752
VERSION
BC012752.1 GI:15215322
KEYWORDS
human.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1793)
Strausberg,R.
Direct Submission
Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 24 Row: m Column: 19.
Location/Qualifiers
1. 1793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3623466"
/tissue_type="Pancreas, adenocarcinoma"
/clone_lib="NIH_MGC_39"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
<1..1219
/codon_start=2
/product="Similar to transmembrane protease, serine 4"
/protein_id="AAH12752.1"
/db_xref="GI:15215323"
/translation="ARGIIALLSLASIIIVLVILKILVDKYYFLCGQLHFIPRKQL

CDS
source
```

```
CDGELDCPLGDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTFA
LATACRQMYSKSPFRAVEIDGDDQDVLVEITEITNESQELMRNRSQPCLSGSLSLH
CLACGSLKTPRVVGGEEASVSWPMQVSIQYDKQHVCGGSIIDPHWLVFAAHCFRKH
TDVFNKVRAGSDKLGSPFLAVAKIIIEFNPMYKPDNDIALMQLQFPFTFSQVVRP
ICLPFDEELTPATPLWIIIGWGTQKNGKMSDILLQASVQVLDKSTCNADDAVQGEV
TEKMKCAGIPEGGVDTCCQSDSGPLMYQSDQHVVGIVSWGVCGGPSTFQVTKVSA
YLNWIIYNVKAEI"
BASE COUNT 436 a 501 c 486 g 370 t
ORIGIN

Query Match 11.1%; Score 193.2; DB 9; Length 1793;
Best Local Similarity 55.8%; Pred. No. 4.4e-27;
Matches 435; Conservative 0; Mismatches 333; Indels 12; Gaps 3;

Qy 924 tctctccctcagtggtcccaactgcgactgagggccatgacggcgagatcggtggaggg 983
Db 471 TCTCCCTGCATGTCCTTGCCTGTGGGAAGAGCCTGAAGACCCCGCTGTGGTGGGTGGG 530
Qy 984 cgctggcctcggatagcaagtgccctggcaagtgtgacttcgacacttcggcaccaccaca 1043
Db 531 AGGAGGCCCTCTGTGGATTCCTTGGCCTTGGCAGGTCAGCATCCAGTACGACAAACAGCAG 590
Qy 1044 tctgtgagacacgctcattgacgcccagtggtgtgtcactgcgcccactgcttctcg 1103
Db 591 TCTGTGGAGGAGCATCTCTGGACCCCGCCACTGGGTGCTTCAACGGCAGCCACTGTCTTC 646
Qy 1104 tgaccccgggagaaggtcctgtgaggggtggaaggtgtacgcgggcaccagcaactcacc 1163
Db 647 --AGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCAGACAAACTGGGCA 704
Qy 1164 agttgctgagggcagcctccacttgccgagatcatatcaacgacattacaccgatgag 1223
Db 705 GCTTCCC---ATCCCTGGCTGTGGCCAAAGATCATCATCTGAATTCACACCCCATGTACC 761
Qy 1224 agagcactatgacatgcctcatgcgctgtcccaagccctgacccctgctcagctcaca 1283
Db 762 CCAAGACAAATGACATCGCCCTCATGAAGTGCAGTGTCCACTTCATCTTCTTCAGGGCAG 821
Qy 1284 tcacacctgttgcctcccccatgcatgagacacctttagcctcaatgagacctgtgga 1343
Db 822 TCAGGCCCATCTCTCTGCCCTCTTTGATGAGGAGGCTCACTCCAGCCACCCCACTCTGGA 881
Qy 1344 tcacaggtcttgcaagaccaggagacagatgacaagacatccccctctccctcgggag 1403
Db 882 TCATTGGATGGGGCTTTACGAAGACAAATGGAGGGAAGATGTCTGACATCTACTGTGACGG 941
Qy 1404 tgcaggtcaatctcatcgacttcaagaaatgcaatgactacttggctctatgacagtacc 1463
Db 942 CGTCAGTCCAGTCAATTGACAGCACACGCTGCAATGCAGACCATGGGTACCAAGGGGAAG 1001
Qy 1464 ttaccccaaggatgatgtgtgtgctggtggaccttcgtggggcagagagactctgccaggag 1523
Db 1002 TCACCGAGAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGCCAGGTG 1061
Qy 1524 acagcggggggcctcttctgtgtgagcagaacacccgtgtacctggtcaggtgtcaca 1583
Db 1062 ACAGTGGTGGGGCCCTGTATGT---ACCAATCTGACACAGTGGCATGTGTGGGCATCTGTTA 1118
Qy 1584 gctggggcacaggtgtgtggccagagaaacacacctgggtgtgtacacaaagtgcagaga 1643
Db 1119 GCTGGGGTATGGCTGCGGGGGCCCGAGCACCCCGAGGAGTATACACCAAGGTCTCAGCCT 1178
Qy 1644 tcttccctgatttacagcaagatgagagcgaggtgtcgattcagaaaaatcctaaccag 1703
Db 1179 ATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTGTAATGCTGCTGCCCTTTGCAG 1238

RESULT 14
AR142620
LOCUS
DEFINITION
Sequence 18 from patent US 6203979.
ACCESSION
AR142620
linear PAT 08-AUG-2001
```

```

VERSION      AR142620.1  GI:15103906
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    Unclassified.
AUTHORS      1 (bases 1 to 2038)
              Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,
              Tang,Y.Tom and Shah,P.
TITLE        Human protease molecules
JOURNAL      Patent: US 6203979-A 18 20-MAR-2001;
FEATURES     Location/Qualifiers
              source
                1..2038
                /organism="unknown"
BASE COUNT   462 a 591 c 569 g 416 t
ORIGIN
Query Match      11.1%; Score 193.2; DB 6; Length 2038;
Best Local Similarity 55.8%; Pred. No. 4.4e-27;
Matches 435; Conservative 0; Mismatches 333; Indels 12; Gaps 3;

Qy  924  tctccctccagtgtcccaactgcggaactgagggccatgacccgagcggtatcgtggagggg 983
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  759  TCTCCCTGCACGTGCTTCCTCTGTGGGAGAGCGCTGAAGACCCCGCTGTGGTGGTGGG 818
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  984  cgtggcctcgatagcaagtggccttgccaagtgaagtctgcaacttcggcaccaccaca 1043
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  819  AGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGTCAGCATCCAGTACGACAAACAGCAG 878
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1044  tctgtgaggcaacgctcaattgacgcccagtggtgctcaactgcgcccacactgctcttcg 1103
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  879  TCTGTGGAGGAGCATCTGTGACCCCGCTGAGTGGTCTTCCAGCGCAGCCCACTGCTTC---- 934
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1104  tgaccgggagaaggtcctgagggctggaaggtgtacgagggcagcagcaactgcacc 1163
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  935  --AGGAACATACCGATGTTTCAACTGGAAGGTGCGGGCAGGCTCAGACAACTGGGCA 992
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1164  agttgctgagcagcctccattgcccagatcatcatacaacagcaattacaccgatgagg 1223
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  993  GCTTCCC---ATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCCCATGTACC 1049
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1224  aggaacactatgacatgcccctcatgctggtgttccaaagccctgaccctgtccgctcaca 1283
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1050  CCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTCTCAGGCACAG 1109
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1284  tccaccctgctgctcccccagatgacagaccctttagacctcaatgagacctgctaga 1343
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1110  TCAGGCCCATCTGCTGCTCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCACTCTGGA 1169
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1344  tcacagggtttggcaagaccaggagacagatgacaagacatcccccttccctccggagg 1403
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1170  TCATTGGATGGGCTTTTACGAAGCAGATGGAGGGAAGATGCTGACATACTGCTGCAGG 1229
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1404  tgcagggtcaatctcatgactcaagaataatgcaatgactacttgcttatgacagttacc 1463
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1230  CGTCAGTCCAGGTATTGACACACACACCGGTGCAATGCAATGCAAGATGCGTACCAGGGGAAG 1289
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1464  ttacccaagatgaatgtgctgggaccttcgtggggcagagactcctgcaggag 1523
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1290  TCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGGTGTGGACACCTGCCAGGGTG 1349
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1524  acagcgggggccccttctgtgtgagcagaacaacccgctggttacctggcaggtgtcaca 1583
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1350  ACAGTGGTGGGCCCTGTATGT---ACCAATCTGACCACTGGGCATGCTGTGGGCATCGTTA 1406
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1584  gctggggcagaggtgtggccagagaacaacctggtgtgtacaccaagtgcacagaag 1643
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1407  GCTGGGGCTATGCTGCGGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTCAGCCT 1466
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1644  ttcttccctgatttacaagaagtggagagcgaggtgcgattcagaaaaatcctaaccag 1703
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1467  ATCTCAACTGGATCTACAATCTCTGGAAGGCTGAGCTGAATGTGCTGTGCCCTTTTGACG 1526
      ||||| || || || || || || || || || || || || || || || || || || || ||
```

```

RESULT 15
AX092380      2063 bp      DNA      linear      PAT 21-MAR-2001
LOCUS        Sequence 111 from Patent WO0116318.
ACCESSION    AX092380
VERSION      AX092380.1  GI:13444504
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 2063)
AUTHORS      Eaton,D.L., Pilvaroff,E., Gerritsen,M.E., Goddard,A.,
              Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
              Wood,W.I.
TITLE        Secreted and transmembrane polypeptides and nucleic acids encoding
              the same
JOURNAL      Patent: WO 0116318-A 111 08-MAR-2001;
              Genentech, Inc. (US)
FEATURES     Location/Qualifiers
              source
                1..2063
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT   477 a 591 c 576 g 419 t
ORIGIN
Query Match      11.1%; Score 193.2; DB 6; Length 2063;
Best Local Similarity 55.8%; Pred. No. 4.4e-27;
Matches 435; Conservative 0; Mismatches 333; Indels 12; Gaps 3;

Qy  924  tctccctccagtgtcccaactgcggaactgagggccatgacccgagcggtatcgtggagggg 983
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  760  TCTCCCTGCACGTGCTTCCTCTGTGGAGAGCGCTGAAGACCCCGCTGTGGTGGTGGG 819
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  984  cgtggcctcgatagcaagtggccttgccaagtgaagtctgcaacttcggcaccaccaca 1043
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  820  AGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGTCAGCATCCAGTACGACAAACAGCAG 879
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1044  tctgtgaggcaacgctcaattgacgcccagtggtgctcaactgcgcccacactgctcttcg 1103
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  880  TCTGTGGAGGAGCATCTGTGACCCCGCTGAGTGGTCTTCCAGCGCAGCCCACTGCTTC---- 935
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1104  tgaccgggagaaggtcctgagggctggaaggtgtacgagggcagcagcaactgcacc 1163
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  936  --AGGAACATACCGATGTTTCAACTGGAAGGTGCGGGCAGGCTCAGACAACTGGGCA 993
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1164  agttgctgagcagcctccattgcccagatcatcatacaacagcaattacaccgatgagg 1223
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  994  GCTTCCC---ATCCCTGGCTGTGGCCAAAGATCATCATATTGAATTCACCCCATGTACC 1050
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1224  aggaacactatgacatgcccctcatgctggtgttccaaagccctgaccctgtccgctcaca 1283
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1051  CCAAGACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACTTCTCAGGCACAG 1110
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1284  tccaccctgctgctcccccagatgacagaccctttagacctcaatgagacctgctaga 1343
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1111  TCAGGCCCATCTGCTGCTCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCACTCTGGA 1170
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1344  tcacagggtttggcaagaccaggagacagatgacaagacatcccccttccctccggagg 1403
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1171  TCATTGGATGGGCTTTTACGAAGCAGATGGAGGGAAGATGCTGACATACTGCTGCAGG 1230
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1404  tgcagggtcaatctcatgactcaagaataatgcaatgactacttgcttatgacagttacc 1463
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1231  CGTCAGTCCAGGTATTGACACACACACCGGTGCAATGCAATGCAAGATGCGTACCAGGGGAAG 1290
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1464  ttacccaagatgaatgtgctgggaccttcgtggggcagagactcctgcaggag 1523
      ||||| || || || || || || || || || || || || || || || || || || || ||
Db  1291  TCACCGAGAAGATGATGTGTGAGGCATCCCGGAAGGGGGTGTGGACACCTGCCAGGGTG 1350
      ||||| || || || || || || || || || || || || || || || || || || || ||

Qy  1524  acagcgggggccccttctgtgtgagcagaacaacccgctggttacctggcaggtgtcaca 1583
      ||||| || || || || || || || || || || || || || || || || || || || ||
```

[illegible]

Search completed: August 19, 2002, 18:39:56
Job time: 9434 sec

OM of: US-09-879-792-12 to: GenEmbl.* out_format : pfs
Date: Aug 20, 2002 5:38 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xip
-O=/cgn2.1/USPTO.spool/US09879792/runat_15082002.104128.22236/app_query.fasta_1.629
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -MODALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODAL=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09879792@cgn1.1.8012
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-879-792-12
Query length: 562
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 2599.010000

score_list:

Sequence	Strd	Orig	zscore	EScore	Len	Documentation
gb_pat:AX149579	+	2991.00	1516.78	4.0e-76	1689	! AX149579 Sequence 3 from Patent
gb_pat:AX149581	+	2963.00	1501.42	2.9e-75	2067	! AX149581 Sequence 5 from Patent
gb_pr:AB048796	+	2956.00	1496.94	5.1e-75	2393	! AB048796 Homo sapiens msp1 mRNA
gb_pr:AB048797	+	2775.50	1404.42	7.3e-70	3324	! AB048797 Homo sapiens msp1 mRNA
gb_pat:AX190543	+	2287.00	1166.17	1.4e-56	1314	! AX190543 Sequence 14 from Patent
gb_pat:AX190546	+	2287.00	1166.17	1.4e-56	1314	! AX190546 Sequence 17 from Patent
gb_pr:AK027798	+	2207.00	1122.73	3.6e-54	2192	! AK027798 Homo sapiens cDNA FLJ
gb_pat:AX190545	+	1869.00	958.33	5.1e-45	1078	! AX190545 Sequence 16 from Patent
gb_ro:BC010843	+	925.00	481.43	1.9e-18	2125	! BC010843 Mus musculus, similar
gb_ro:AF243500	+	790.50	415.48	8.8e-15	1735	! AF243500 Mus musculus plasmid
gb_ro:AF199362	+	786.50	413.40	1.2e-14	1755	! AF199362 Mus musculus serine H
gb_ro:AF113596	+	785.50	412.91	1.2e-14	1753	! AF113596 Mus musculus mosaic s
gb_ro:AB073550	+	763.50	403.06	4.3e-14	1473	! AB073550 Rattus norvegicus mRNA
gb_ro:MM0300738	+	758.00	396.14	1.1e-13	2752	! MM0300738 Mus musculus mRNA for
gb_pat:AX354825	+	744.00	388.67	2.8e-13	2348	! AX354825 Sequence 1 from Patent
gb_pat:AX201156	+	743.00	387.53	3.2e-13	3245	! AX201156 Sequence 786 from Patent
gb_pat:AX267955	+	743.00	387.53	3.2e-13	3245	! AX267955 Sequence 929 from Patent
gb_pat:AX201158	+	741.50	392.04	1.8e-13	1476	! AX201158 Sequence 788 from Patent
gb_pat:AX267957	+	741.50	392.04	1.8e-13	1476	! AX267957 Sequence 931 from Patent
gb_pat:AX201157	+	741.50	392.02	1.8e-13	1479	! AX201157 Sequence 787 from Patent
gb_pat:AX267956	+	741.50	392.02	1.8e-13	1479	! AX267956 Sequence 930 from Patent
gb_pat:AX207907	+	740.50	391.27	2.0e-13	1656	! AX207907 Sequence 11 from Patent
gb_pr:AF270487	+	740.50	390.44	2.2e-13	1740	! AF270487 Homo sapiens androgen
gb_pr:AF329454	+	739.00	385.57	4.1e-13	3226	! AF329454 Homo sapiens epitheli
gb_pr:AF123453	+	738.50	390.52	2.2e-13	1479	! AF123453 Homo sapiens transmem
gb_pr:AB038157	+	736.50	386.11	3.8e-13	2468	! AB038157 Homo sapiens TMRSS3a
gb_pat:AX201121	+	736.50	386.08	3.8e-13	2479	! AX201121 Sequence 751 from Patent
gb_pat:AX207965	+	736.50	386.08	3.8e-13	2479	! AX207965 Sequence 69 from Patent
gb_pat:AX267920	+	736.50	386.08	3.8e-13	2479	! AX267920 Sequence 894 from Patent
gb_pat:AX306771	+	736.50	386.08	3.8e-13	2479	! AX306771 Sequence 4 from Patent
gb_pr:HS075329	+	736.50	386.08	3.8e-13	2479	! HS075329 Human serine protease m
gb_pat:AX041973	+	736.50	382.94	5.7e-13	3966	! AX041973 Sequence 3 from Patent
gb_pr:AB038158	+	731.00	383.12	5.6e-13	2554	! AB038158 Homo sapiens TMRSS3b
gb_pat:AX354827	+	729.50	386.58	3.6e-13	1359	! AX354827 Sequence 3 from Patent
gb_pat:AX277411	+	713.50	375.56	1.5e-12	2135	! AX277411 Sequence 69 from Patent
gb_pr:AB038159	+	713.50	375.56	1.5e-12	2135	! AB038159 Homo sapiens TMRSS3c
gb_pat:AX207967	+	713.00	375.49	1.5e-12	2079	! AX207967 Sequence 71 from Patent
gb_pr:AF216312	+	713.00	375.49	1.5e-12	2079	! AF216312 Homo sapiens type II
gb_ov:AB038497	+	712.50	371.75	2.4e-12	3507	! AB038497 Xenopus laevis Hesp-2
gb_pr:AF179224	+	710.50	374.23	1.8e-12	2081	! AF179224 Homo sapiens transmem

gb_pr:BC011703 + 710.50 374.21 1.8e-12 2088 ! BC011703 Homo sapiens, Simi
gb_pat:AR142620 + 707.50 372.87 2.1e-12 2038 ! AR142620 Sequence 18 from p
gb_pat:AX092380 + 707.50 372.54 2.2e-12 2063 ! AX092380 Sequence 111 from p
gb_pat:AX207899 + 701.50 369.55 3.2e-12 2137 ! AX207899 Sequence 3 from Pa
gb_pat:AX076190 + 697.50 370.84 2.7e-12 1305 ! AX076190 Sequence 1 from Pa
seq_name: gb_pat:AX149579

seq_documentation block:

LOCUS AX149579 1689 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 3 from Patent WO0136604.
ACCESSION AX149579
VERSION AX149579.1 GI:14348013
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1689)
AUTHORS Madison, E. I., and Ong, E. O.
TITLE Nucleic acids encoding endothelases, endothelases and uses thereof
JOURNAL Patent: WO 0136604-A 3 25-MAY-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source
1..1689
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1689
/note="DNA encoding a transmembrane serine protease (Endothelase 2-S) protein"
/codon_start=1
/protein_id="CAC41220.1"
/db_xref="GI:14348014"

BASE COUNT 367 a 529 c 470 g 323 t
ORIGIN
alignment_scores:
Quality: 2991.00 Length: 562
Ratio: 5.332 Gaps: 0
Percent Similarity: 99.822 Percent Identity: 99.822

alignment_block:

US-09-879-792-12 x AX149579 ..
Align seg 1/1 to: AX149579 from: 1 to: 1689
1 MetCluArgaspSerHisGlyAsnAlaSerProAlaargThrProSerAl 17
|||||
1 ATGGAGAGGAGGAGCCAGCGGATGTCATCCAGCAAGACACCTTCAGG 50
|||||
17 aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArga 34
|||||
51 TGGAGCATCTCCAGCCAGGCATCTCCAGTGGGACACCTCCAGCCGGG 100
|||||
34 laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
|||||
101 CATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGTGGGACACCT 150
|||||
51 ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProGly 67
|||||
151 CCGGGCCGGGCGATCTCCAGCCAGGCATCTCCAGTGGTACACCTCCAGG 200

```
67 yArgAlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaArgA 84
|||||
201 CCGGGCATCTCCAGCGCGGGCATCTCCAGCCAGGCATCTCCAGCCCGG 250
|||||
84 laSerProAlaLeuAlaSerLeuSerArgSerSerGlyArgSerSer 100
|||||
251 CATCTCCGGCTCTCGGCATCACTTTCCAGGTCTCATCCGGCAGGTCATCA 300
|||||
101 SerAlaArgSerAlaSerValThrThrSerProThrArgValThrLeuVa 117
|||||
301 TCCGCCAGGTCTCGGTGACAACTTCCCAACACAGAGTGTACCTTGT 350
|||||
117 lArgAlaThrProValGlyAlaValProIleArgSerSerProAlaArgS 134
|||||
351 TAGAGCAACACCACTGGGGGTGTACCATCCGATCATCTCCTGCCAGGT 400
|||||
134 erAlaProAlaThrArgAlaThrArgGluSerProGlyThrSerLeuPro 150
|||||
401 CAGCACCAGCAACAGGGGCCACCAAGGAGAGCCAGGTACGAGCTGCC 450
|||||
151 LysPheThrTrpArgGluGlyGlnLysGlnLeuProLeuIleGlyCysVa 167
|||||
451 AAGTTCACTGGCGGAGGCCAGAACGACTACCGCTCATCTGGGTGGT 500
|||||
167 lLeuLeuLeuIleAlaLeuValValSerLeuLeuIleLeuPheGlnPheT 184
|||||
501 GCTCCTCCTATTGCCCTGGTGGTTCGGCTCATCTCCTTCCAGTTCT 550
|||||
184 rpGlnGlyHisThrGlyIleArgThrLysGluGlnArgGluSerCysPro 200
|||||
551 GGCAGGGCCACAGAGGATCAGGTACAAAGGAGCAGAGGGAGAGCTGTCCC 600
|||||
201 LysHisAlaValArgCysAspGlyValValAspCysLysLysSerAs 217
|||||
601 AAGCACGCTGTTCCCTCTGACGGGGTGGTGAAGTCAAGCTGAAGAGTGA 650
|||||
217 pGluLeuGlyCysValArgPheAspTrpAspLysSerLeuLysIleT 234
|||||
651 CGAGCTGGGTGCTGAGTGGTGTGACTGGGCAAGTCTCTGCTTAAATCT 700
|||||
234 yrSerGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsn 250
|||||
701 ACTCTGGGTCTCCCATCAGTCAGTGGCTTCCCATCTGTACGACCACTGGAA 750
|||||
251 AspSerTrpSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHi 267
|||||
751 GACTCTTACTCAGAGAAGACCTGCAGCAGCTGGGTTTCGAGAGTGTCTCA 800
|||||
267 sArgThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleL 284
|||||
801 CCGGACAACCGAGGTGCCACAGGGATTTGGCCAAACAGCTTCTCAATCT 850
|||||
284 euArgTrpAsnSerThrIleGlnGluSerLeuHisArgSerGluCysPro 300
|||||
851 TGAGATACACTCCACCATCCAGGAAGACCTCCACAGGTCTGAATGCCCT 900
|||||
301 SerGlnArgTrpIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMe 317
|||||
901 TCCAGCGGTATATCTCCCTCCAGTGTTCACACTGCGGACTGAGGGCCAT 950
|||||
317 tThrGlyArgIleValGlyAlaLeuAlaSerAspSerLysTrpProT 334
|||||
951 GACCGGCGGATCGTGGGAGGGCGCTGGGCTCGGATAGCAAGTGGCCTT 1000
|||||
334 rpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeu 350
|||||
1001 GCGAAGTGAGTCTGCATCTCGGCACCAACCCACATCTGTGAGGACGCTC 1050
|||||
351 IleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhePheValThrAr 367
|||||
1051 ATTGACCCCGAGTGGGTGCTCACTGCGGCCCACTGTCTTCTCGTAGCCG 1100
|||||
```

```
367 gGluLysValLeuGluGlyTrpLysValTyrAlaGlyThrSerAsnLeuH 384
|||||
1101 GGAGAAGGTCTCGAGGCTGGAAGGTGTACGGCGGCACGACCAACCTGC 1150
|||||
384 isGlnLeuProGluAlaAlaSerIleAlaGluIleIleIleAsnSerAsn 400
|||||
1151 ACCAGTTGCTTGAGGCGCTCCATTCGCGAGATCATCATCAACAGCAAT 1200
|||||
401 TyrThrAspGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLy 417
|||||
1201 TACACCGATGAGGAGGACGACTATGACATGCCCTCATGGCTGTCCAA 1250
|||||
417 sProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProMethHisG 434
|||||
1251 GCCCTGACCTGTCCGCTCACATCCACCTGCTTGCCTCCCATGTCATG 1300
|||||
434 lyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLys 450
|||||
1301 GACAGACCTTTAGCCTCAATGAGACCTGTGGATCACAGGCTTTGGCAAG 1350
|||||
451 ThrArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnVa 467
|||||
1351 ACCAGGAGACAGATGACAAGACATCCCTTCTCCGGGAGGTGCAGGT 1400
|||||
467 lAsnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrAspSerT 484
|||||
1401 CAATCTCATGACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGTT 1450
|||||
484 yrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAsp 500
|||||
1451 ACCTTACCCCAAGGATGATGTGCTGGGACCTTCGTGGGGCGAGAGAC 1500
|||||
501 serCysGlnGlyAspSerGlyGlyProLeuValCysGluGlnAsnAsnAr 517
|||||
1501 TCCTGCCAGGAGACACGCGGGGCGCTTGTCTGTGAGCAGACAACCG 1550
|||||
517 gTrpTyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgA 534
|||||
1551 CTGGTACTGGCAGGTGTACCCAGCTGGGCGACAGGCTGGGCCAGAGAA 1600
|||||
534 snLysProGlyValTyrThrLysValThrGluValLeuProTrpIleTyr 550
|||||
1601 ACAACCTGGTGTGTACACCAAGTGACAGAGTTCTTCCCTGGATTTC 1650
|||||
551 SerLysMetGluSerGluValArgPheArgLysSer 562
|||||
1651 AGCAAGATGGAGCGAGGTGCGGATTCATATAAAATCC 1686
|||||
```

seq_name: gb_pat:AX149581

seq_documentation_block:

LOCUS AX149581 2067 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent WO0136604.

ACCESSION AX149581

VERSION AX149581.1 GI:14348015

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2067)

AUTHORS Madison,E.I. and Ong,E.O.

TITLE Nucleic acids encoding endotheliases, endotheliases and uses thereof

JOURNAL Patent: WO 0136604-A 5 25-MAY-2001;

CORVAS INTERNATIONAL, INC. (US)

FEATURES

source

1..2067 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..2067

/note="DNA sequence encoding a transmembrane serine

protease (endotheliase-L) protein"

```
/codon_start=1
/protein_id="CAC41221.1"
/db_xref="GI:14348016"
/translation="MERDSHGNSPARTPSAGASPAQASPACTPPGRASPAQASPAQA
SPAGNPGRASPAQASPACTPPGRASPAQASPAQASPAQASPAQASPAQASPAQAS
RSASVTSFTRVLYRATPVGAVPIRSSPARSAPATRAITRSPGTSLSKFTWREGQKQ
LPLGCVLLIALVSLILQFQWGHGIRYKEQRESCPHAVRCDGVDCVCKLKSDE
LGCVRFDKLLKTIYSSGSHWLPICSSNNDYSSEKTCQOLGFESAHRTTEVAHRD
FANFSILRYNSTIOESLHRECPQSORYISQCSHGRLRAMTGRIVGGALASDSKWPW
QVSLHFTGTHICGGTLLDAQWVLTAAHCFEFTREKVLGKWKVYAGTSLNHLQPEASI
AEILINSNTDEEDDYDALMELSKPLTSAHIHPACLPMHGOTFSLNETCWTGFGK
TREDTDXTPLEQVQVNLIDFKCNDLYLVDSYLTPRMCGADLRRGDRSCQGSQG
FLVCGNNRWLAGVSWGTGCGQRKQGVYKTVETVLPFIYSKMNRAQRVERKAWTY
RPGQLLRGCRSPRSIFLCKVAMDFENVSAPDEFIVFVIKHLCMGIRSSWPPALFY
LVFLFFLLLLLSFLKNTSDSILTLTFTAVTRMLPENYHSFPFPLKIFHONLTIK
EYKVIK"
```

BASE COUNT 463 a 604 c 540 g 460 t
ORIGIN

alignment_scores:
Quality: 2963.00 Length: 558
Ratio: 5.310 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.283

alignment_block:
us-09-879-792-12 x AX149581 ..

Align seg 1/1 to: AX149581 from: 1 to: 2067

```
1 MetGluArgAspSerHisGlyAsnAlaSerProAlaAlaArgThrProSerAl 17
|||||
1 ATGGAGAGGGACAGCCACGGGAATGCATCTCCAGCAAGAACACCTTCAGC 50
17 aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgA 34
|||||
51 TGGAGCATCTCCAGCGCGGACATCTCCAGTGGGACACCTCCAGCGCGGG 100
34 laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
|||||
101 CATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGCTGGGACACCT 150
51 ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProGly 67
|||||
151 CCGGGCGGGCATCTCCAGCCAGGCATCTCCAGCTGGTACACCTCCAGG 200
67 yArgAlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaArgA 84
|||||
201 CCGGGCATCTCCAGCGCGGACATCTCCAGCCAGGCATCTCCAGCGCGGG 250
84 laSerProAlaLeuAlaSerLeuSerArgSerSerGlyArgSerSer 100
|||||
251 CATCTCCGGCTCTGGGCATCATCTTCAGGTCTTCATCCCGCAGGTATCA 300
101 SerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeuVa 117
|||||
301 TCGCGCAGGTCAGCTCGGTGACAACTCCCAACACAGAGTGTACCTTGT 350
117 laArgAlaThrProValGlyAlaValProLleArgSerSerProAlaArgS 134
|||||
351 TAGAGCAACACCATGGGGGCTGTACCCATCCGATCATCTCCCTGCCAGGT 400
134 erAlaProAlaThrArgAlaThrArgGluSerProGlyThrSerLeuPro 150
|||||
401 CAGCACCAACCAACAGGCCACCGAGGAGGACCGAGGTACGAGCCCTGCC 450
151 LysPheThrTrpArgGluGlyGlnLysGlnLeuProLeuIleGlyCysVa 167
|||||
451 AAGTTACCTTGGCGGAGGCGGACAGAGCAGTACCCTCATCTCGGTGGGT 500
167 lleuLeuLeuIleAlaLeuValValSerLeuIleIleLeuPheGlnPhe 184
|||||
501 GCTCCCTCCTCATTTGCCCTGGTGGTTTTCGTCATCATCTCTCCAGTTCT 550
```

```
184 rpGlnGlyHisThrGlyIleArgTyrLysGluGlnArgGluSerCysPro 200
|||||
551 GGCAGGCCACACAGGGATCAGGTACAGGAGCAGGAGGAGCTGTCCC 600
201 LysHisAlaValArgCysAspGlyValValAspCysLysLeuLysSerAs 217
|||||
601 AAGCACCTCTTCGCTGTGACGGGGTGTGACTGCAAGCTGCAAGAGTGA 650
217 pGluLeuGlyCysValArgPheAspTrpAspLysSerLeuLysIleT 234
|||||
651 CGAGCTGGGTGCTGTGAGGTTTGACTGGGACAGCTCTGCTTAAATCT 700
234 yrSerGlySerSerHisGlnTrpLeuProLleCysSerSerAsnTrpAsn 250
|||||
701 ACTCTGGGTCTCCCATCAGTGGCTTCCCATCTGTAGCAGCAACTGGAAT 750
251 AspSerTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHi 267
|||||
751 GACTCTACTCAGAGAAGACTGCCAGCAGCTGGGTTTCGAGAGTGCTCA 800
267 sArgThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleL 284
|||||
801 CCGGACAAACGAGGTTCGCCACAGGGATTTGCCACAGCTTCTCAATCT 850
284 euArgTyrAsnSerThrIleGlnGluSerLeuHisArgSerGluCysPro 300
|||||
851 TGAGATACAACCTCCACCATCCAGGAAAGCTCCACAGCTCTGAATGCCCT 900
301 SerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMe 317
|||||
901 TCCAGCGGTATATCTCCCTCCAGTGTTCCTACTGCGGACTGAGGGCAT 950
317 tThrGlyArgIleValGlyAlaLeuAlaSerAspSerLysTrpProt 334
|||||
951 GACCGGGCGGATCTGGGAGGGCGCTGGGCTCGGATAGCAAGTGGCCTT 1000
334 rpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeu 350
|||||
1001 GGCAAGTGAGTGTCACTTCGGCACCAACCCACATCTCTGGAGGCACGCTC 1050
351 IleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhePheValThrAr 367
|||||
1051 ATTCAGCCCGAGTGGGTGCTCACTGCGCGCCACTGCTTCTTCGTGACCCG 1100
367 gGluLysValLeuGluGlyTrpLysValTyrAlaGlyThrSerAsnLeuH 384
|||||
1101 GGAGAGGTCTGGAGGGCTGGAGGTGTACGGGGCACCAGCAACCTGC 1150
384 isGlnLeuProGluAlaAlaSerIleAlaGluIleIleIleAsnSerAsn 400
|||||
1151 ACCAGTTGCCCTGAGGACGCTCCATTCGCCGAGATCATCATCAACAGCAAT 1200
401 TyrThrAspGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLy 417
|||||
1201 TACACCGATGAGGAGGACGATATGACATCGCCCTCATCGGGTGTCCAA 1250
417 sProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProMethiSg 434
|||||
1251 GCCTCTGACCTGTCCGCTCACATCCACCTGCTTGCTTCCCTCCCATGATG 1300
434 lyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLys 450
|||||
1301 GACAGACCTTTAGCTCAATGAGACCTGCTGGATCAGAGGCTTTGGCAG 1350
451 ThrArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnVa 467
|||||
1351 ACCAGGGAGACAGATGACAAAGACATCCCTTCTCCGGGAGGTGCAAGT 1400
467 laSnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrAspSerT 484
|||||
1401 CAATCTCATCGACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGTT 1450
```



```

Quality: 2775.50          Length: 572
Ratio: 5.277             Gaps: 2
Percent Similarity: 91.958  Percent Identity: 91.958

alignment_block:
US-09-879-792-12 x AB048797 ..

Align seg 1/1 to: AB048797 from: 1 to: 3324

1  MetGluArgAspSerHisGlyAsnAlaSerProAlaArgThrProSerAl 17
88 ATGGAGAGGACACCCACGGGATGCATCTCCAGCAAGAACACCTTCAGC 137
17  aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProProGlyArg 34
138 TGGAGCATCTCCAGCCAGGACATCTCCAGCTGGGACACCTCCAGGCCGG 187
34  laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
188 CATCTCCAGGCCACGGCATCTCCAGCCAGGACATCTCCAGCTGGGACACCT 237
51  ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProProGl 67
238 CCGGCCGGGCATCTCCAGCCAGGACATCTCCAGCTGGTACACCTCCAGG 287
67  yArgAlaSerProGlyArg.....A 74
288 CCGGGCATCTCCAGCCGGGCGGCATCTCCAGCCAGGACATCTCCAGCCCGCAG 337
74  laSerProAlaGlnAlaSerProAlaArgAlaSerProAlaLeuAlaSer 90
338 CATCTCCAGGCCACGGCATCTCCAGCCGGGCGGCATCTCCAGCTGGGACATCA 387
91  LeuSerArgSerSerGlyArgSerSerAlaArgSerAlaSerVa 107
388 CTITCCAGGTCCTCATCCGCGAGGTCTATCCGCCAGGTCAGCCTCGGT 437
107 lThrThrSerProThrArgValTyrLeuValArgAlaThrProValGlyA 124
438 GACAACCTCCCAACCAAGAGGTACCTTGTAGAGCAACACCAAGTGGGG 487
124 laValProIleArgSerSerProAlaArgSerAlaProAlaThrArgAla 140
488 CTGTACCCATCCGATCATCTCTCCAGGTCAGACCAACCAAGGCCC 537
141 ThrArgLysSerProGlyThrSerLeuProLysPheThrTrpArgGluG 157
538 ACCAGGGAGAGCCCA..... 552
157 yGlnLysGlnLeuProLeuIleGlyCysValLeuLeuLeuIleAlaLeuV 174
552 ..... 552
174 alValSerLeuIleLeuPheGlnPheThrGlnGlyHisThrGlyIle 190
553 .....GTCCAGTTCTGGCAGGGCCACACAGGGATC 582
191 ArgTyrLysGluGlnArgGluSerCysProLysHisAlaValArgCysAs 207
583 AGGTACAAGGAGGAGAGGAGCTGTCCCAAGCACGCTGTTCGCTGTGA 632
207 pGlyValValAspCysLysLeuLysSerAspGluLeuGlyCysValArgp 224
633 CCGGGGTGGTGTGACTGCAAGCTGAAGAGTACAGAGCTGGGCTGGCTGAGGT 682
224 heAspTrpAspLysSerLeuLeuLysIleTyrSerGlySerHisGln 240
683 TTGACTGGGCAAGTCTCTGCTTAAATCTACTCTGGGCTCTCCCATCAG 732
241 TrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrSerGluLysTh 257
733 TGGCTTCCCATCTGTAGCACCACTGGAATGACTCTCTACTCTACAGAGAAGAC 782
```

```

257 rCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAlaH 274
783 CTGCCAGCAGCTGGGTTTCGAGAGTGTCTACCGGACAACGAGGTTGCC 832
274 isArgAspPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrIle 290
833 ACAGGGATTTGGCAACAGCTTCTCAATCTTGAGATACAACACTCCACCATC 882
291 GlnGluSerLeuHisArgSerGluCysProSerGlnArgTyrIleSerLe 307
883 CAGGAACGCTCCACAGGTCTGAATGCCCTTCCACGCGGTATATCTCCCT 932
307 uGlnCysSerHisCysGlyLeuArgAlaMetThrGlyArgIleValGlyG 324
933 CCAGTGTTCCTCACTGCGGACTGAGGGCCATGACCGGCGGATCGTGGGAG 982
324 lyAlaLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPhe 340
983 GGGCGCTGGCTCGGATAGCAAGTGGCTTGGCAAGTGAGTCTGCACCTTC 1032
341 GlyThrThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLe 357
1033 GGCACCAACCCACATCTCTGGAGGCACGCTCATTCAGCCCGAGTGGGTGCT 1082
357 uThrAlaAlaHisCysPhePheValThrArgGluLysValLeuGluGlyT 374
1083 CACTGCCGCCCACTGCTTCTTCGTACCGGGAGAGGTCCTGGAGGGCT 1132
374 rPlysValTyrAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAla 390
1133 GGAAGGTGTACGGGGCACCAACCTGCACCAGTTGCCCTGAGGCAGCC 1182
391 SerIleAlaGluIleIleAsnSerAsnTyrThrAspGluGluAspAs 407
1183 TCCATTGCCGAGATCATCAACAGCAATATACACCGATGAGGAGGACGA 1232
407 pTyrAspIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaH 424
1233 CTATGACATCGCCCTCATGGGCTGTCCAAGCCCTGACCTGTCCGCTC 1282
424 isIleHisProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsn 440
1283 ACATCCACCTGCTTGCTCCCATGTCATGTCAGACAGACCTTTAGCTCAAT 1332
441 GluThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLy 457
1333 GAGACCTGCTGGATCACAGGCTTTGGCAAGCACCGGAGACAGATGACAA 1382
457 sThrSerProPheLeuArgGluValGlnValAsnLeuIleAspPheLysL 474
1383 GACATCCCCCTTCTCCGGGAGGTGCGAGGTCAATCTCATCGACTTCAAGA 1432
474 yCysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMet 490
1433 AATGCAATGACTACTTGGTCTATGACAGTTACCTTACCCCAAGGATGATG 1482
491 CysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGl 507
1483 TGTGCTGGGACCTTCTGTTGGGGCAGAGACTCTCTGCCAGGGAGACAGCGG 1532
507 yGlyProLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValT 524
1533 GGGGCTCTTGTCTGTGAGCAGAACACCGCTGTACCTGGCAGGTGTCA 1582
524 hrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThr 540
1583 CCAGCTGGGGCACAGGCTGTGGCCAGAGAAACAAACCTGGTGTGTACACC 1632
541 LysValThrGluValLeuProTrpIleTyrSerLysMetGluSerGluVa 557
1633 AAAGTGACAGAAAGTCTTCTCCCTGGATTATACAGCAAGATGAGGAGCGAGGT 1682
557 lArgPheArgLysSer 562
```

```
|||||
1683 GCGATTGAGAAATCC 1698
seq_name: gb_pat:AX190543
seq documentation block:
LOCUS AX190543 1314 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 14 from Patent WO0136645.
ACCESSION AX190543
VERSION AX190543.1 GI:15143854
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1314)
AUTHORS Quinn,K.E., Spytak,K.A., Majumder,K., Vernet,C., Burgess,C.,
Fernandes,B., Taupier,R.J., Rastelli,L. and Herrmann,J.L.
TITLE Serine/threonine kinase and serine protease polypeptides and
nucleic acids encoding same
JOURNAL Patent: WO 0136645-A 14 25-MAY-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
1..1314
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1266
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC50789.1"
/db_xref="GI:15143855"
/translation="MESPTSLPKFTWREGQKQLPLIGCVLLILIALVVSLLILFQFW
GHTGRYREORESCPKHVRCDGVDDKLSDELGCVRFDMDKSLKIVSGSSHOWLP
ICSSNWDPSYSEKTCQQLGFESAHTTEVAHRDPANSPILRYNSTIQESLHRSQPS
QRYSLSQSHGLRAMTRIGVGGALASDKWPQVSLHFGTHICGGHLDIQWVITA
AHCFTVREKVLGVKWKVYAGTSNLHQLPEASIAEIIINSNTDEEDYDIALMLRSK
PLTUSALHPLACLPMHGGTFSLNETCWTGFKTRTDKTSPLFREYQVNLIDFKKC
NDLYVDSYLTFRNMKAGDLRGDRSCQGGSGGPLVCBQNNRWYLAGVTSWGTGCGQR
NKPGVYTKVTEVLPWIYSKMESEVRFTKS"
BASE COUNT 288 a 384 c 376 g 266 t
ORIGIN
alignment_scores:
Quality: 2287.00 Length: 420
Ratio: 5.458 Gaps: 0
Percent Similarity: 99.762 Percent Identity: 99.762
alignment_block:
US-09-879-792-12 x AX190543 ..
Align seg 1/1 to: AX190543 from: 1 to: 1314
143 GluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGlnLy 159
|||||
4 GAGAGCCAGGTACGAGCCTGCCAAAGTTCCACCTGGCGGGAGGCCAGAA 53
|||||
159 sGlnLeuProLeuIleGlyCysValLeuLeuLeuIleAlaLeuValVal 176
|||||
54 GCAGCTACCGCTCATCGGGTGCCTGCTCCCTCCATGTCCTCGCTGGTGT 103
|||||
176 erLeuIleLeuPheGlnPheTrpGlnGlyHisThrGlyIleArgTyr 192
|||||
104 CGCTCATCCTCTTCCAGTTCTTGGCAGGGCCACACAGGGATCAGGTAC 153
|||||
193 LysGluGlnArgGluSerCysProLysHisAlaValArgCysAspGlyVa 209
|||||
154 AAGGAGCAGAGGAGAGCTGTCCCAAGACGCTGTTGGCTGTGACGGGT 203
|||||
209 lValAspCysLysLeuLysSerAspGluLeuGlyCysValArgPheAsp 226
|||||
204 GGTGGACTGCAGCTGAAGAGTACGAGCTGGCTGCGTGAGGTTTGACT 253
```

```
226 rpAspLysSerLeuLeuLysIleTyrSerGlySerSerHisGlnTrpLeu 242
|||||
254 GGGACAAGTCTCTGCTTAAAATCTACTCTGGGTCTCTCCCATCAGTGGCTT 303
|||||
243 ProIleCysSerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGl 259
|||||
304 CCCATCTGTAGCACAACAGTGAATGACTCTCTACTCAGAGAAGACCTGCCA 353
|||||
259 nGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAlaHisArgA 276
|||||
354 GCAGCTGGGTTTCGAGAGTGCTCACCGGACACCGAGGTGGCCACAGGG 403
|||||
276 spPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrIleGlnGlu 292
|||||
404 ATTTTGCCAAACAGCTTCTCAATCTTGAGATACAACTCCACCATCCAGAA 453
|||||
293 SerLeuHisArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCy 309
|||||
454 AGCCTCCACAGGTCTGAATGCCCTTCCCAGCGGTATATCTCCCTCCAGTG 503
|||||
309 sSerHisCysGlyLeuArgAlaMetThrGlyArgIleValGlyGlyAlaL 326
|||||
504 TTCCCACTGGGACTGAGGGCCATGACCGGGCGGATCGTGGGAGGGGCGC 553
|||||
326 euAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThr 342
|||||
554 TGGCTCGGATAGCAAGTGGCTTTGGCAAGTAGTCTGCACCTTCGGCACC 603
|||||
343 ThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAl 359
|||||
604 ACCCACATCTGTGGAGCAGCGCTCATTTGACGCCAGTGGGTGCTCACTGC 653
|||||
359 aAlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLysV 376
|||||
654 CGCCACTGCTCTCTGTCGCCGGGAGAGGCTCTGAGGGCTGGAGAGG 703
|||||
376 alTyrAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIle 392
|||||
704 TGTACGGCGGCACCCAGCAACCTGCACCACTTGCTGAGGACGCCCTCCATT 753
|||||
393 AlaGluIleIleIleAsnSerAsnTyrThrAspGluIleAspAspTyrAs 409
|||||
754 GCGGAGATCATCATCAACAGCAATTAACCCGATGAGGAGGACGACTATGA 803
|||||
409 pIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleH 426
|||||
804 CATCGCCTCATCGGGCTGTCCAAGCCCCCTGACCCCTGTCGCTCACATCC 853
|||||
426 isProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsnGluThr 442
|||||
854 ACCTGCTTGCCTCCCATGCATGGACAGACCTTTAGCCTCAATGAGACC 903
|||||
443 CysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLysThrSe 459
|||||
904 TGCTGGATCAGAGCTTTGGCAAGACCCAGGAGACAGATGACAAGACATC 953
|||||
459 rProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysCysA 476
|||||
954 CCCCTTCTCCGGGAGGTGCGAGGTCAATCTCATCGACTTCAAGANAATGCA 1003
|||||
476 snAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAla 492
|||||
1004 ATGACTACTTGGTCTATGACAGTTACCTTACCCCAAGGATGATGTGTGCT 1053
|||||
493 GlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyPr 509
|||||
1054 GGGAGACCTTCGTGGGGCAGAGACTCCTGCCAGGAGACAGCGGGGGGCC 1103
|||||
509 oLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSerT 526
|||||
1104 TCTGTCTGTGACGAGACAACCCGCTGGTACCTGGCAGAGGTGTACACAGCT 1153
|||||
526 rpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysVal 542
```

```

|||||
1154 GGGCCACAGGCTGGCCAGAGAAACAAACCTGGTGTGTACACCAAGTG 1203
|||||
543 ThrGluValLeuProTrpIleTyrSerLysMetGluSerGluValArgPh 559
|||||
1204 ACAGAAGTTCTTCCCTGGATTACAGCAAGATGAGAGCGAGGTGGCATT 1253
|||||
559 eArgLysSer 562
| | | | |
1254 CACAAATCC 1263

seq_name: gb_pat:AX190546
seq_documentation_block:
LOCUS AX190546 1314 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 17 from Patent WO0136645.
ACCESSION AX190546
VERSION AX190546.1 GI:15143857
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1314)
AUTHORS Quinn,K.E., Spytek,K.A., Majumder,K., Vernet,C., Burgess,C.,
Fernandes,E., Taupier,R.J., Rastelli,L. and Herrmann,J.L.
TITLE Serine/threonine kinase and serine protease polypeptides and
nucleic acids encoding same
JOURNAL Patent: WO 0136645-A 17 25-MAY-2001;
Curagen Corporation (US)
FEATURES
source
location/Qualifiers
1..1314
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 266 a 376 c 384 g 288 t
ORIGIN

alignment_scores:
Quality: 2287.00 Length: 420
Ratio: 5.458 Gaps: 0
Percent Similarity: 99.762 Percent Identity: 99.762

alignment_block:
US-09-879-792-12 x AX190546/rev ..
Align seg 1/1 to reverse of: AX190546 from: 1 to: 1314

143 GluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGlnLy 159
|||||
1311 GAGAGCCAGGTACGAGCCTGCCCAAGTTCACTGGCGGAGGGCCAGAA 1262
|||||
159 sGlnLeuProLeuIleGlyCysValLeuLeuLeuIleAlaLeuValVal 176
1261 GCAGTACCGCTCATCGGTGGTGGTCTCTCTCCATTTGCCCTGGTGGTTT 1212
|||||
176 erLeuIleLeuPheGlnPheTrpGlnGlyHisThrGlyIleArgTyr 192
|||||
1211 CGCTCATCATCTCTTCCAGTTCTGGCAGGCGCCACACAGGGATCAGGTAC 1162
|||||
193 LysGluGlnArgGluSerCysProLysHisAlaValArgCysAspGlyVa 209
|||||
1161 AAGGACAGAGGGAGAGCTGTCCCAACACGCTGTTCCGTGTGACGGGGT 1112
|||||
209 lValAspCysLysLeuLysSerAspGluLeuGlyCysValArgPheAspT 226
|||||
1111 GGTGGACTGAAGCTGAAGAGTGACGAGCTGGGCTGGCTGAGGTTTGACT 1062
|||||
226 rpAspLysSerLeuLeuLysIleTyrSerGlySerSerHisGlnTrpLeu 242
|||||
1061 GGGCAAGTCTCTGCTTAAATCTACTCTGGGTCTCCCATCATGTCGCTT 1012
|||||
243 ProLecysSerSerAsnTrpAsnAspSerTyrSerGluLysThrCysG1 259
|||||

```

```

|||||
1011 CCCATCTGTAGCAGCAACTGGAATGACTCTCTACTCAGAGAGACCTGCCA 962
|||||
259 nGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAlaHisArgA 276
|||||
961 GCAGCTGGTTTCGAGAGTGCTCACCGGACAACCGAGGTTGCCACAGGG 912
|||||
276 spPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrIleGlnGlu 292
|||||
911 ATTTTGCCAAACAGCTTCTCAATCTTGAGATACAACTCCACCATCCAGAA 862
|||||
293 SerLeuHisArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCy 309
|||||
861 AGCTCCACAGGCTGAATGCCCTTCCAGCGGATATCTCCCTCCAGTG 812
|||||
309 sSerHisCysGlyLeuArgAlaMetThrGlyArgIleValGlyGlyAla 326
|||||
811 TTCCCACTGCGGACTGAGGCCATGACCGGCGGATCGTGGAGGGCGC 762
|||||
326 euAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThr 342
|||||
761 TGGCCTCGGATAGCAAGTGGCCTTGGCAAGTGAGTCTGCACTTGGGCACC 712
|||||
343 ThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAl 359
|||||
711 ACCCACATCTGTGGAGCACGCTCATTTGACGCCAGTGGGTGCTCACTGC 662
|||||
359 aAlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLysV 376
|||||
661 CGCCCACTGCTTCTTCGTGACCGGGAGAGGTCCTGGAGGGCTGGAAGG 612
|||||
376 alTyrAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIle 392
|||||
611 TGTACGGGGCACCAAGCACTGCACCAAGTTGCTGAGGAGCCCTCCATT 562
|||||
393 AlaGluIleIleAsnSerAsnTyrThrAspGluGluAspAspTyrAs 409
|||||
561 GCGGATCATCATCAACAGCAATTACCCGATGAGGAGGACGACTATGA 512
|||||
409 pIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleH 426
|||||
511 CATCGCCCTCATCGGGCTGTCCAAGCCCTGACCTGTCGCTCACATCC 462
|||||
426 isProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsnGluThr 442
|||||
461 ACCCTGCTTGCCTCCCATGATGACAGACACTTTAGCCCTCAATGAGACC 412
|||||
443 CysTrpIleThrGlyPheGlyLysThrArgGluThrAspLysThrSe 459
|||||
411 TGCTGGATCACAGGCTTTGCCAAGACCAGGAGACAGATCACAGACATC 362
|||||
459 rProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysCysA 476
|||||
361 CCCTTCTCTCGGAGGTGCGAGGTCAATCTCATCGACTTCAAGAAATGCA 312
|||||
476 snAspTyrLeuValTyrAspSerTyrIleThrProArgMetMetCysAla 492
|||||
311 ATGACTACTTGGTCTATGACAGTTACCTTACCCCAAGGATGATGTGCT 262
|||||
493 GlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyPr 509
|||||
261 GGGGACCTTCTGGGGCAGAGACTCTCTGCCAGGAGACAGCGGGGGGCC 212
|||||
509 oLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSert 526
|||||
211 TCTTGTCTGTGAGCAGAAACAACCGCTGGTACCTGGCAGGTGTCAACAGCT 162
|||||
526 rpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysVal 542
|||||
161 GGGGCACAGGCTGTGGCCAGAGAAACAACCTGGTGTGTACACCAAGTG 112
|||||
543 ThrGluValLeuProTrpIleTyrSerLysMetGluSerGluValArgPh 559
|||||

```


941 ACAGCTTCTCAATCTTGAGATACAACTCCACCATCCAGGAAGCCTCCAC 990
296 ArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCysSerHisCy 312
991 AGGTCTGAATGCCCTTCCAGCGGTATATCTCCCTCCAGTGTCCCACTG 1040
312 sGlyLeuArgAlaMetThrGlyArgIleValGlyGlyAlaLeuAlaSerA 329
1041 CGGACTGAGGCGCCATGACCGGGCGGATCGTGGGAGGGCGCTGGCCCTCG 1090
329 spSerLysTyrProTyrPrlnValSerLeuHisPheGlyThrThrHisIle 345
1091 ATAGCAAGTGGCCCTGGCAAGTGAAGTCTGCACCTCGGCACCCACCATC 1140
346 CysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCy 362
1141 TGTGAGGCGCAGCTCATTTAGCGCCAGTGGGTGCTCACTGCGGCCCACTG 1190
362 sPhePheValThrArgGluLysValLeuGluGlyTyrPrlsValTyrAlaG 379
1191 CTTCCTGTCGACCGCGGAGAGGTCCTGGAGGGCTGGAAGGTGTACGCGG 1240
379 lyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIleAlaGluIle 395
1241 GCACACGAACCTGCACAGTGTGCTGAGGCGCTCCATGTCGCGAGATC 1290
396 IleIleAsnSerAsnTyrThrAspGluGluAspTyrAspIleAlaLe 412
1291 ATCATCAACAGCAATTACACCGATGAGGAGGACGACTATGACATCGCCCT 1340
412 uMetArgLeuSerLysProLeuThrLeuSerAlaHis..... 424
1341 CATCGGCTGTCCAAAGCCCTGACCCCTGTCGGTGGAGGAATCTGCACCT 1390
425 ..IleHisProAlaCysLeuProMethis 433
1391 CCCGCTCTCTGCCCCCAGCCCGCAGCAC 1419

seq_name: gb_pat:AX190545

seq_documentation_block:
LOCUS AX190545 1078 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 16 from Patent WO0136645.
ACCESSION AX190545
VERSION AX190545.1 GI:15143856
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1078)
AUTHORS Quinn,K.E., Spytek,K.A., Majumder,K., Vernet,C., Burgess,C.,
Fernandes,E., Taupier,R.J., Rastelli,L. and Herimann,J.L.
TITLE Serine/threonine kinase and serine protease polypeptides and
nucleic acids encoding same
JOURNAL Patent: WO 0136645-A 16 25-MAY-2001;
Curagen Corporation (US)

FEATURES
source
1..1078
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 221 a 296 c 319 g 242 t
ORIGIN

alignment_scores:
Quality: 1869.00 Length: 342
Ratio: 5.481 Gaps: 0
Percent Similarity: 99.708 Percent Identity: 99.708

alignment_block:
US-09-879-792-12 x AX190545/rev ..

Align seg 1/1 to reverse of: AX190545 from: 1 to: 1078
221 CysValArgPheAspTyrAspLysSerLeuLeuLysIleTyrSerGlySe 237
1077 TCGGTGAGGTTTGTACTGGGACAAGTCTCTCTTAAAAATCTACTCTGGGTC 1028
237 rSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrS 254
1027 CTCCCATCAGTGGCTTCCCATCTGTAGCAGCAACTGGAATGACTCTCTACT 978
254 erGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThr 270
977 CAGAGAAGACCTGCCAGCAGCTGGGTTTCGAGAGTGTCTACCGGACAACC 928
271 GluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrAs 287
927 GAGGTTGCCACACAGGATTTTGGCAACAGCTTCTCAATCTTGAGATACAA 878
287 nSerThrIleGlnGluSerLeuHisArgSerGluCysProSerGlnArgT 304
877 CTCACCATCCAGGAAAGCCTCCACAGGTCTGAATGCCCTTCCACAGCGGT 828
304 yrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMetThrGlyArg 320
827 ATATCTCCCTCCAGTGTTCACACTGCGGACTGAGGGCCATGACCGGCGG 778
321 IleValGlyGlyAlaLeuAlaSerAspSerLysTyrProTyrPrlnValse 337
777 ATCGTGGGAGGGCGCTGGCTCGGATAGCAAGTGGCTTGGCAAGTGAG 728
337 rLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuIleAspAlaG 354
727 TCTGCACCTTCGGCACCCACCATCTGTGGAGGCACGCTCATTTGACGCC 678
354 lnTrpValLeuThrAlaAlaHisCysPhePheValThrArgGluLysVal 370
677 AGTGGGTGCTCACTGCGGCCACTGCTTCTCGTGACCCCGGAGAGGTC 628
371 LeuGluGlyTyrLysValTyrAlaGlyThrSerAsnLeuHisGlnLeuPr 387
627 CTGGAGGGCTTGGAGGTGTACGGGGGCACCAGCAACCTGCACCAAGTTGCC 578
387 oGluAlaAlaSerIleAlaGluIleIleAlaAsnSerAsnTyrThrAspG 404
577 TGAGGAGAGCCTCCATTCGCCGAGATCATCAACAGCAATTACACCGATG 528
404 luGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysProLeuThr 420
527 AGGAGGAGCAGTATGACATCGCCCTCATGCGGTGTCTCAAGCCCTGACC 478
421 LeuSerAlaHisIleHisProAlaCysLeuProMethisGlyGlnThrPh 437
477 CTGTCCGCTCACATCCACCTGCTTGGCTCCCATGATGATGACAGACCTT 428
437 eSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluT 454
427 TAGCCCAATGAGACCTCTGGATCAGAGCTTGGCAAGACCAGGAGAGA 378
454 hrAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuIle 470
377 CAGATGACAAGACATCCCCCTTCTCCGGGAGGTGCGAGGTCAATCTCATC 328
471 AspPheLysLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrPr 487
327 GACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGTTTACCTTACCCC 278
487 oArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnG 504
277 AAGATGATGTGTCTGGGACCTTCTGTGGGGCAGAGACTCTCTGCCAGG 228
504 lyAspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTyrTrpTyrLeu 520
227 GAGACAGCGGGGGCGCTCTTGTGTGTGTGAGCAGAAACAACCGCTGTACCTG 178

521 AlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProG1 537
 |||||
 177 GCAGGTGTCCACGAGTGGGGCACAGGCTGTGGCCAGAGAAACAACTGG 128
 |||||
 537 yValTyrThrLysValThrGluValLeuProTrpIleTyrSerLysMetG 554
 |||||
 127 TGTGTACACCAAGGTGACAGAAAGTTCTCCCTGGATTACAGCAAGATGG 78
 |||||
 554 luSerGluValArgPheArgLysSer 562
 |||||
 77 AGAGCGAGGTGCGATTCAAAAATCC 52

seq_name: gb_ro:BC010843

seq documentation_block:

LOCUS BC010843 2125 bp mRNA linear ROD 17-JUL-2001
 DEFINITION Mus musculus, Similar to mosaic serine protease, clone
 IMAGE:3490022, mRNA, partial cds.

ACCESSION BC010843

VERSION BC010843.1 GI:14789982

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2125)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (12-JUL-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mdcpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 7 Row: b Column: 18

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

Source

1..2125

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:3490022"

/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months

old, gross tissue."

/clone_lib="NCI_CGAP_Mam5"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

<1..531

/codon_start=1

/product="Similar to mosaic serine protease"

/protein_id="AAH10843.1"

/db_xref="GI:14789983"

/translation="PEAASISQIIINGNYTDEDDYDIALIRLSKPLTSAIHFPACL

PMHGQTFGLNETCTWTFGKTKETKTSFPLREYQVNLIDFKNDYLVYDSYLTFR

MMCADLRGGRSCOGDGGPLVCEQNNRWYLAGVTSWGTGCGQKNKPGVYTKVTEVL

PHYRKMESEVFRKS"

536 a 502 c 541 g 546 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 925.00 Length: 176

Ratio: 5.286 Gaps: 0

Percent Similarity: 99.432 Percent Identity: 94.318

alignment_block:

US-09-879-792-12 x BC010843

Align seg 1/1 to: BC010843 from: 1 to: 2125

387 ProGluAlaAlaSerIleAlaGluIleIleAlaSerAsnTyrThrAs 403

|||||

1 CCCGAGGCTGCCTCTATCTCCAGATCATCATCAACGGCAACTACACAGA 50

403 pGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysProLeuT 420

|||||

51 TGAACAGAGTACTATGACATTGCCCTCATCAGGCTGTCCAAGCCCTGA 100

420 hrLeuSerAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThr 436

|||||

101 CCCTGTCAAGTCACATCCACCTGCCTGCCCTCCCGATGCAGGTCAGACC 150

437 PheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgG1 453

|||||

151 TTCGGCCTCAATGAGACCTGTGGATCACGGCTTTGGCAAAACCAAGGA 200

453 uThrAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuI 470

|||||

201 GACAGATGAGAGACATCTCCCTTCCCGAGAGGTTCCAGGTCAACCTCA 250

470 leAspPheLysLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThr 486

|||||

251 TTGACTTCAAGAAGTGCATCACTACTTGGTCTATGACAGACTACCTTACC 300

487 ProArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAspSerCysG1 503

|||||

301 CCAAGGATGATGTGTGCGGGGATCTACGAGAGGGAGGACTCCTGCGCA 350

503 nGlyAspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTrpTyrL 520

|||||

351 GGGAGACAGTCGGAGACCTCGTCTGTGACAGAACAAATCGCTGGTACC 400

520 euAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysPro 536

|||||

401 TGGCAGGTGTCCACAGCTGGGGCACAGGCTGTGGCCAGAAAACCAAGCT 450

537 GlyValTyrThrLysValThrGluValLeuProTrpIleTyrSerLysMe 553

|||||

451 GGTGTGTACACCAAGTGACAGAAAGTACTTCCCTGGATTACAGAAAGAT 500

553 tGluSerGluValArgPheArgLysSer 562

|||||

501 GGAGAGTGAGGTACGATTCGCGAAATCT 528

seq_name: gb_ro:AF243500

seq documentation_block:

LOCUS AF243500 1735 bp mRNA linear ROD 17-APR-2000

DEFINITION Mus musculus plasmic transmembrane protein X mRNA, complete cds.

ACCESSION AF243500

VERSION AF243500.1 GI:7578918

KEYWORDS

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 1735)

AUTHORS Han, J. and Kim, S. O.

TITLE Putative transmembrane protease X

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1735)

AUTHORS Han, J. and Kim, S.O.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Immunology, The Scripps Research Institute,
10550 N. Torrey Pines Rd., La Jolla, CA 92126, USA

FEATURES
source 1..1735 /organism="Mus musculus"
/db_xref="taxon:10090"
CDS 101..1573 /note="putative protease"
/codon_start=1
/product="plasmic transmembrane protein X"
/protein_id="AAF64186.1"
/db_xref="GI:7578919"
/translation="MALNSGSPGIGPCYENHGYSQHEHCIPRPVPAPNGYNLYPAQY
YPSPVQYAPRITTOASTSVIHTHPKSSGPACTSKKSLSCLALAGTGLVTGAAVAAY
LLWRFDWNSCNRTSEMCGSGGTCTSSSLMCDGVAHCPNGEDENRCVRLYGQSFILOYV
SSQRAWPYQCDDWSESTGRACKDMGYKNKFNNFYSSQIPDQSGATSFMKLNYSSGNV
DYLYKKLHSDSCSRMVLRCECGRVSRIQROSLRVGGNLASPGDWPMOVSILHVGN
HVGCGSIITPEWTVAHCVREPLSGPRYWFATAGLQLSMFYGRSHOVKEKYVISHPN
YDSKTNNNDIALAKLATPLAFNLDLPVKPLPNCPLMDLDOECWISGWGATYEKGKTS
DVLNAWPILEPSKNSKIYYNNLTTPAMICACFLQGSVDSCQSDSGGPLVTLTKNGI
WWLIQDTSWGSGCAKALRPGVCYGNVTFTDIWIQQMRANS"

BASE COUNT 441 a 439 c 457 g 398 t
ORIGIN

alignment_scores:
Quality: 790.50 Length: 568
Ratio: 2.395 Gaps: 15
Percent Similarity: 58.099 Percent Identity: 32.570

alignment_block:
US-09-879-792-12 x AF243500 ..

Align seg 1/1 to: AF243500 from: 1 to: 1735

```

10 SerProAlaArgThrProSerAlaGlyAlaSerProAlaGlnAlaSerPr 26
   ||||| ||||| ||||| : : : : : ||||
62 GCCCATACTGAACCTCCATCTGCTGCACACAGCGGATGGCATTTAA 111
   ||||| ||||| ||||| : : : : : ||||

26 oAlaGlyThrProProGlyArgAlaSerProAlaGlnAlaSerProAlaG 43
   ||||| ||||| ||||| : : : : : ||||
112 CTCAGGGTCACCTCCAGGA...ATCGGACCTTGCTATGAGAACCACGGT 158
   ||||| ||||| ||||| : : : : : ||||

43 InAlaSerProAlaGlyThrProProGlyArgAlaSerProAlaGlnAla 59
   ||||| ||||| ||||| : : : : : ||||
159 ATCAGTCTCAGCACATCTGCTCTCCG..... 184
   ||||| ||||| ||||| : : : : : ||||

60 SerProAlaGlyThrProProGlyArgAlaSerProGlyArgAlaSerPr 76
   ||||| ||||| ||||| : : : : : ||||
185 .....AGACCACACCTGGCTCCCAATGGCTACAACTTGATCC 222
   ||||| ||||| ||||| : : : : : ||||

76 oAlaGlnAlaSerProAlaArgAlaSerProAlaLeuAlaSerLeuSerA 93
   ||||| ||||| ||||| : : : : : ||||
223 ACCCCAGTAATACTACCA..... 238
   ||||| ||||| ||||| : : : : : ||||

93 rgSerSerSerGlyArgSerSerSerAlaArgSerAlaSerValThrThr 109
   ||||| ||||| ||||| : : : : : ||||
238 ..... 238
   ||||| ||||| ||||| : : : : : ||||

110 SerProThrArgValTyrrLeuValArgAlaThrProValGlyAlaValPr 126
   ||||| : : : ||||| ||||| : : : : : ||||
239 TCTCCAGTGCCTCAGTATGCTCCGAGGATTACAACGCCATCAACATC 288
   ||||| ||||| ||||| : : : : : ||||

126 o...ileArgSerSerProAlaArgSer...AlaProAlaThrArgAlaT 141
   ||||| : : : ||||| ||||| : : : : : ||||
289 TGTCATCCACACATATCCCAGTCTCAGGAGCACCGTGCACCTCAAAC 338
   ||||| ||||| ||||| : : : : : ||||

141 hrArgGluSerProGlyThrSerLeuProLysePhetTrpArgGluGly 157
   ||||| : : : ||||| ||||| : : : : : ||||
339 CTAAAGAAATCGGTGTTTTAGCCCTTGCCCTGGGCAT..... 376
   ||||| ||||| ||||| : : : : : ||||

```



```
543 GAACCGTGTGTTTCGTCTCTACAGGACAAAGCTTCATCTCCTCAGGCTTTACT 592
235 erGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAsp 251
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
593 CATCTCAGAGAAAGCCTGGTATCCCGTGTGCCAGGATGATTGGAGTGAG 642
252 SerTyrSerGluTysThrCysGlnGlnLeuGlyPheGluSerAlaHisAr 268
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
643 AGCTACGGAGAGCAGCATGTAAAGACATGGATACAAAGCAACATTTTAA 692
268 gThrThrGlu...ValAlaHisArgAspPheAlaAsnSerPheSerIleL 284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 TTCTAGCCAGGGATACACAGCAGAGCGGGCAACAGCTTTATGAAGC 742
284 euArgTyrAsnSer.....ThrIleGlnGluSerLeuHisArg 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
743 TGAATGTGAGCTCAGGCATGTTGACCTCTATAAAACTCTACCACAGT 792
297 SerGluCysProSerGlnArgTyrIleSerLeuGlnCysSerHisCysG 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
793 GACTCATGTTTCATCCCGCATGGTGTCTTCTTGGCTGTATAGAATCGG 842
313 yLeuArgAlaMet.....ThrGlyArgIleValGlyAlaLeuAlas 328
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
843 GGTTCCGTCTAGTGAACCCAGCAGCAGATGTGGGTGGATGTAATGCCT 892
328 erAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThrThrHis 344
|| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
893 CACCAGGAGACTGSCCCTGGCAGGTGACGTCGACGTCACGAGCGCTCAC 942
345 IleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHi 361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
943 GTCTGCGGAGGCTCCATCATCACCCCGCAGTGGATTGTGACGCGGCCCA 992
361 sCysPhePheValThrArgGluLysValLeuGluGly.....TrpL 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
993 CTGTGTG.....GAAGAACCCCTCAGCGGCCCGAGGTACTGGA 1030
375 ysValTyrAlaGlyThr.....SerAsnLeu 383
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1031 CGGCATTTGCGGAATTCCTCAGACAGCTCTCTCATGTTCCATGGAAGTGA 1080
384 HisGlnLeuProGluAlaAlaSerIleAlaGluIleIleAsnSerAs 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1081 CACCAG.....GTAGAAAAGTAATTTCCCATCCAAA 1112
400 nTyrThrAspGluAspTyrAspIleAlaLeuMetArgLeuSerL 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1113 TTAGGACTCTAAGACCAGAAATACGACATGCTCTCATGAAGCTGCAGA 1162
417 ysProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProMethis 433
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1163 CACCTTTGGCTTTTAATGATCTAGTAGAGCCAGTGTCTCTGCCGAACCCA 1212
434 GlyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLy 450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1213 GCGATGATGTAGACCTAGACCAAGAAATGCTGGATTTCGGGGTGGGGGGC 1262
450 sThrArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnV 467
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1263 CACCTATGAG...AAGGGAAGACCTCGGACGTGTTGAATGCTGCCATGG 1309
467 alAsnLeuIleAspPheLysCysAsnAspTyrLeuValTyrAspSer 483
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1310 TACCCTTGATCGAGCCCTCCAAATGTAATAGTAAATACATATACAAAC 1359
484 TyrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyGlyArg 500
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1360 CTAATCACACAGCCATGATCTGTGCCGGCTTCTCCAGGGGTCTGTCTGA 1409
500 pSerCysGlnClyAspSerGlyGlyProLeuValCysGluGlnAsnAsnA 517
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1410 CTCCTGCCAGGACAGCTGGAGGCGCGCTGGTTACTTTCGAAGAATGGGA 1459
```

```
517 rgntTyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArg 533
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1460 TCTGGTGGCTGATTGGGACACAGAGCTGGGCTCGGCTGTGCCAAGGCA 1509
534 AsnLysProGlyValTyrThrLysValThrGluValLeuProTrpIleTy 550
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1510 CTCAGACCTGAGTATACGGGAGGTGACGGTATTACAGATTGGATCTA 1559
```

550 rSerLysMetGluSer 555

1560 CCAGCAATGAGGCG 1575

seq_name: gb_ro:AF113596

seq_documentation_block:

LOCUS AF113596 1753 bp mRNA linear ROD 06-MAR-2000

DEFINITION Mus musculus mosaic serine protease epitheliasin mRNA, complete

ACCESSION AF113596 cds.

VERSION AF113596.1 GI:6648959

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1753)

AUTHORS Jacquinet,E., Rao,N.V., Rao,G.V. and Hoidal,J.R.

TITLE Cloning, genomic organization, chromosomal assignment and

expression of a novel mosaic serine proteinase: epitheliasin

FEBS Lett. 468 (1), 93-100 (2000)

JOURNAL 20148617

MEDLINE 2 (bases 1 to 1753)

REFERENCE Jacquinet,E.J., Rao,N.V., Rao,G.N. and Hoidal,J.R.

AUTHORS Direct Submission

TITLE Submitted (14-DEC-1998) Pulmonary Division, University of Utah, 50

JOURNAL North Medical Drive, Salt Lake City, UT 84132, USA

FEATURES Location/Qualifiers

1..1753

/organism="Mus musculus"

/strain="BALB/c"

/db_xref="taxon:10090"

/chromosome="16"

/map="telomeric region of long arm"

101..1573

/codon_start=1

/product="mosaic serine protease epitheliasin"

/protein_id="AAF21308.1"

/db_xref="GI:6648960"

/translation="MALNSGSPPGIGPCYENHGYOSEHICPPRPVAPNGYNLYPAQY

YPSVPQYAPRITTOASTSVIHTHPKSSGAPCTSKSKSLCLALALGTVLTGAATAAV

LILREWDSNCSTSEMECSLGTICISSLWCDGVAHCPNGEDENRCVRLYGOSFILOY

SSQRKAWYPVQDDMSYNGRAACKDMGYKNFVSSQIGPDQSGATSPFKLNVSSGNV

DLYKKLHSDSCSRMVSRLRCIECGVRSVKRQSRIVGGLNASPGDMPQVSLHVQGV

HVCGGSIITPEWITVAACHVEPRLTAFAGILRQSLMFYSGRHRQVKEVISHPN

YDSKTKNDIALLKLTPLAENDLYKPLCPNPGMMLDLDQECWISGHWGATYEKGTKS

DVLNAAMPLEIPSRKNSKIYNNLITPAMICAGFLQGSVDSCQSDSGGPIVTLKNGI

WWLIGDTSWGGCALKALPGVYVTVTFDWTIIQQMRANS"

BASE COUNT 461 a 438 c 455 g 399 t

ORIGIN

alignment_scores:

Quality: 785.50 Length: 568

Ratio: 2.380 Gaps: 15

Percent Similarity: 58.099 Percent Identity: 32.218

alignment_block:

US-09-879-792-12 x AF113596 ..

Align seg 1/1 to: AF113596 from: 1 to: 1753

10 SerProAlaArgThrProSerAlaGlyAlaSerProAlaGlnAlaSerPr 26

[illegible]

```

794 TCCGCGATGGTGGTTCCTTCCTCGCGCTGTATAGAAATCGCGGGTTCGCTCAGT 843
317 t.....ThrGlyArgIleValGlyGlyAlaLeuAlaSerAspSerLysT 332
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::
844 GAAACGCCACGACGAGATGTTGGGTGGATTCGAATGCCTCCACGAGACT 893
332 rpProTrpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGly 348
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
894 GGCCCTGGCAGCTCAGCTGCAGCTCCAAAGCGGTCCAGCTCTCGCGAGGC 943
349 ThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhePheVa 365
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::
944 TCCATCATCACCCCGAGTGGATTGTGACGCCGCCACCTGTGTG..... 988
365 lThrArgGlnLysValLeuGluGly.....TrpLysValTyrAlaG 379
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
989 .....GAAGAACCCTCAGCGGCCGAGTACTGACGGCATTCGCGGTCGGG 1031
379 lYThr.....SerAsnLeuHisGlnLeuPro 387
      ||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1032 GAATTCCTGAGACAGCTCTCATGTTCTATGGAAGTAGACACACG..... 1075
388 GluAlaAlaSerIleAlaGluIleIleAsnSerAsnTyrThrAspG 404
      :::|||||:::|||||:::|||||:::|||||:::|||||:::
1076 .....GTAGAAAAAGTAATTTCCTCATCCAAATTACGACTCTAA 1113
404 uGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysProLeuThrL 421
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1114 GACCAAGAAATACGACATGCTCTCATGAGCTCGACACCTTTGGCTT 1163
421 euSerAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThrPhe 437
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1164 TTAATGATCTAGTGAAGCCAGTGTCTTCGCCGAACCCAGGCATGATGCTA 1213
438 SerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluTh 454
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1214 GACCTAGACCAAGGAATGCTGGATTTCGGGGTGGGGGCCACCTATGAG.. 1261
454 rAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuIleA 471
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1262 ..AAGGGGAAGACCTCGGAGCTGTTGAATGCTGCCATGGTACCCCTTGATCG 1310
471 spPheLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrPro 487
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1311 AGCCCTCCAAATGTAATAGTAATAATACATACAAACACCTTAACACACCA 1360
488 ArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnG 504
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1361 GCCATGATCTGCGGGCTTCCTCCAGGGGCTCTGCAGCTCTTGCCAGGG 1410
504 yaspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTrpTyrLeuA 521
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1411 AGACAGTGGAGGGCGCTGGTTACTTTGAAGAATGGGATCTGTTGGCTGA 1460
521 laGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGly 537
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1461 TTGGGGACACAGCTGGGGCTGCGGGCTGTGCCAAGGACCTCAGACCTGGA 1510
538 valTyrThrLysValThrGluValLeuProTrpIleTyrSerLysMetG 554
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1511 GTATACGGGAACGTGACGGTATTTTACAGATTGGATCTACCAAGCAAAATGAG 1560
554 user 555
      :::
1561 GGCG 1564

_name: gb_ro:AB073550
documentation_block:
US AB073550 1473 bp mRNA linear
INR1 Rattus norvegicus mRNA for TMPRSS2, complete cds.
SSION AB073550
AB073550.1 GI:16444961

```

KEYWORDS
SOURCE Rattus norvegicus cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1473)
AUTHORS Tsuzuki,S.
TITLE TMPRSS2, Rat
JOURNAL Published Only in Database (2001) In press
REFERENCE 2 (bases 1 to 1473)
AUTHORS Tsuzuki,S.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Satoshi Tsuzuki, Kyoto University,
Laboratory of Nutrition Chemistry, Division of Food Science and
Biotechnology, Graduate School of Agriculture; Olwake-cho,
Kitashirakawa, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail:tkmone@kais.kyoto-u.ac.jp, Tel:81-75-753-6263,
Fax:81-75-753-6264)
FEATURES
Location/Qualifiers
1..1473
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..1473
/gene="TMPRSS2"
1..1473
/gene="TMPRSS2"
/codon_start=1
/product="TMPRSS2"
/protein_id="BAB70683.1"
/db_xref="GI:16444962"
1..1473
/translation="MALNSGSPPGIGPYVENHGYQSEHVYSPRPVSPGYNLYPAQS
CSPVOYAPRVYTOASTPAIHQIPSSGTCTCTSKSKMLVALALGLSPVAAVAAG
LWKKWDKSCSSEMEWECGSGGNCISSILMCDGVAQCPNGKDNCRVLYGTSTPTLVY
SQRAWYFVQDDWNEGCRACKDMYKNSEFYSSQGIPIQDSGATSPMKLVNVSAGV
DIYKLYHSDSCSRHWVSILRCEGVRVRSRIVGGSTASPGDWPQWVSLHVQGI
HVCYSITPEWIVTARHCEVEPLSPRYWTAFAGILKQSLMPYGSRHQVEKVISHPNI
YDSTKNNIDALMLQTLAFINDVWRPVCLPNPMLDLAQECWISGWTAEKRGKTS
DVLNAAVPLIEPSKNSRYIYNLLITPAMICAGFLQGSVDSCQSGSGGLPVLTKKEI
WWLIGTWSGSGCAKAYRPGVYGVNVTVDWIYQOMRANS"
BASE COUNT 364 a 380 C 397 G 332 T
ORIGIN
alignment_scores:
Quality: 763.50 Length: 551
Ratio: 2.386 Gaps: 16
Percent Similarity: 58.076 Percent Identity: 33.031
alignment_block:
US-09-879-792-12 x AB073550 ..
Align seg 1/1 to: AB073550 from: 1 to: 1473
27 AlaGlyThrProGlyArgAlaSerProAlaGlnAlaSerProAlaG1 43
:::||||:|||||
13 TCAGGTCACTCCAGGA..... 30
43 nAlaSerProAlaGlyThrProGlyArgAlaSerProAlaGlnAlas 60
:::||||
31 .ATTGGACTTACTATGAGACCCAGCGGTATCAGTCTGAGCAGCTTAT 79
:::||||
60 erProAlaGlyThrProGlyArgAlaSerProGlyArgAlaSerPro 76
||||| ||||| |||
80 CCCC.....AGGCCACTGTGCTCCCACTGCTACACTTGATATCCA 123
|||
77 AlaGlnAlaSerProAlaArgAlaSerProAlaLeuAlaSerLeuSerAr 93
|||||::: |||
124 GCCCAGACTGCCCA..... 138
93 gSerSerSerGlyArgSerSerAlaArgSerAlaSerValThrThrS 110
|||
139T 139


```
388 luAlaSerIleAlaGluIleIleAsnSerAsnTyrThrAspGlu 404
976 .....GTAGAAAAGTGATTCCCATCCAAATACGACTCTAAG 1014
405 GluAspAspTyrAspIleAlaLeuMetArgLeuSerIysProLeuThrIle 421
1015 ACCAAGAAATAATACATGCTCTATGAAGCTGCAGACACCCCTGGCTTT 1064
421 uSerAlaHisIleProAlaCysLeuProMetHisGlyGlnThrPheS 438
1065 TAATGATGATAGTGAACCGACTGTCTGCCGAACCCAGGATGCTGG 1114
438 erLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluThr 454
1115 ACCTAGCCAGGAGTCTGTGATTTCAGGGTGGGGGCCACCTATGAG... 1161
455 AspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuIleAs 471
1162 AAGGGAAGACCTCAGATGTGCTGAATGTCGCATGGTACCCCTTGATCGA 1211
471 pPheLysLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrProA 488
1212 GCCTCCAAATGTAATAGTAATAATACATATACAACTTAATCACACCAG 1261
488 rgMetMetCysAlaGlyAspLeuArgGlyClyArgAspSerCysGlnGly 504
1262 CCATGATCTCGCGGGTCTCTCAGGGTCTGTGACTCTTGTCCAGGGA 1311
505 AspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTrpTyrLeuAl 521
1312 GACAGTGGAGGGCCCTGGTGTACTTTCAGAAATGAATCTGGTGGCTGAT 1361
521 aglyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyV 538
1362 TGGGACACAGAGTGGGCTCGGGCTGTGCCAAGGCATACAGACCTGGAG 1411
538 alTyrThrLysValThrGluValLeuProTrpIleTyrSerLysMetGlu 554
1412 TATACGGGAATGTGACAGATTTTACAGATTGGATCTACCAGCAATGAGG 1461
555 Ser 555
1462 GCG 1464
```

```
seq_name: gb_ro:MMU300738
seq_documentation_block:
LOCUS MMU300738 2752 bp mRNA linear ROD 10-DEC-2001
DEFINITION Mus musculus mRNA for transmembrane serine protease 3 (Tmprs3
gene).
ACCESSION AJ300738
VERSION AJ300738.1 GI:17529622
KEYWORDS Tmprs3 gene; transmembrane serine protease 3.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Guipponi M., Scamuffa N., Scott, H.S., Rossier, C. and
Antonarakis, S.E.
REFERENCE
Isolation and characterization of the mouse Tmprs3 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2752)
AUTHORS Guipponi M.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2001) Guipponi M., Division of Medical Genetics,
University of Geneva Medical School, 1 rue Michel Servet, 1211
Geneva 4, SWITZERLAND
FEATURES
Location/Qualifiers
source 1. 2752
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="17"
```

```
gene 56. .1417
/ gene="Tmprs3"
CDS 56. .1417
/ gene="Tmprs3"
/ function="transmembrane serine protease"
/ codon_start=1
/ product="Tmprs3 protein"
/ protein_id="CAC83350.1"
/ db_xref="GI:17529623"
/ translation="MGNDPPAAEPFSSRLFGLDLKLISPVAPDGDVAVAOILSL
PLKFPPIIVIGIILALALGIGHFDCSKYKCHSFKCIELTARCDGVSDCKNAE
DEYRCRVSGQRAAHQVFTAAARWTMSDDMKSHYAKTACQALGFPSTVSSDHLKRDV
LEEQFGDFVSINHLSDDKVYALHHSVMREGCTSGHVTLKCSACGRTGYSPRIV
GNNSSLQWPMQVSLQFGYHLCGSIITPLWIVTAARCVYDLYHPKSWTVQVGLVS
LMDSPVSHLVEKIIYHSHKPKRLGNDIALMKLSEPLTFDQIICLPNSENFPD
KLCWTSGWGTEDGGDASPVLNHAAVPLISNKICNHRDVGIIISPSMLCAGYLKGG
VDSGQDSGGPLVQERRLKLKLVGATSGIGCAEVNKPQVYTRITSLFDWIHQELER
LKT"
BASE COUNT 636 a 722 c 743 g 651 t
ORIGIN
alignment_scores:
Quality: 758.00 Length: 485
Ratio: 2.445 Gaps: 17
Percent Similarity: 63.918 Percent Identity: 35.670
alignment_block:
US-09-879-792-12 x MMU300738 ..
Align seg 1/1 to: MMU300738 from: 1 to: 2752
125 ValProIleArgSerSerProAlaArgSerAlaPro.....AlaThrAr 139
8 GTGAGGTGGAGCGCGGACCAACATCAGCGCGCTGAGATGACCAT 57
139 gAlaThrArgGluSerProGlyThrSerLeuProLysPheThrTrpArg. 155
58 GGGGGAGAATGACCTCTCGGAGCGGAGTCC...TTCTCCTTCGGT 104
156 .....GluGly 157
105 CTCTCTTCGGCTGGATGATTGAAATAAGTCTCTGTGGCACCAGATGGA 154
158 Gln.....LysGlnLeuProLeu..... 163
155 GACGCAGTGCCTGCGCAGATCTTGTCCTCTGCTTGAAGTTTTC 204
164 .....IleGlyCysValLeuLeuLeuAlaLeuValValSerL 177
205 GATCATCGTCTATGGGATCATCGCTTGATCTGCGCTGGCCATCGGTC 254
177 eutlelleLeuPheGlnPhe.....TrpGlnGlyHisThrGlyIle 190
255 TGGGCATCCACTCGACTGCTCTGGGAAGTACAGGTGCCATTCTTC 304
191 ArgTyrLysGluGlnArgGluSerCysProLysHisAlaValArgCysAs 207
305 AAG.....TGATTGAACCTACACGCTCGATCGGA 333
207 pGlyValValAspCysLysLeuLysSerAspGluLeuGlyCysValArgp 224
334 CGGGGTCTCGACTGCAAGAACGACAGAGGATGATACCGATGCGTGGGG 383
224 heAspTrpAspLysSerLeuLeuLysIleTyrSerGlySerSerHisGln 240
384 TGAGCGGCCCAAGAGCGGCGCACCAAGTGTTCACAGCTGCTGCG... 427
241 TrpLeuProIleCysSerSerAsnTrpAsnAspSerSerGluLysTh 257
428 TGGAGGACCATGTGCTCCGATGATGAAAGTCACTATCCCAAGATTGC 477
257 rCysGlnGlnLeuGlyPhe.....GluSerAlaHisArgThrt 270
```

```
478 CTGTGCCAGCTGGGGTTTCCCACTATGTGAGCTCAGACCACCTAAAGAG 527
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 hrGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyr 286
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
528 TGGATGCGTGGAGGACAGTTCAGGCGCAGCTC.....GTGTCCATC 571
287 AsnSerThrIleGlnGlu.....SerLeuHis..... 295
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
572 AATCACCCTCTGTGCAGATGACAAAGGTGACGCGGTACACCACCTCTGTGTA 621
296 ....ArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCysSerH 311
      ||| :||| ||| :||| :||| :||| :||| :||| :|||
622 CATGAGGGAAGGTGCACCTCAGCCACCGCTGCTACCTTGAAGTCTCAG 671
311 isCysGlyLeuArgAlaMetThrGly.....ArgIleValGlyGly 324
      ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
672 CCGTGGCAGCAGA.....ACGGCTACAGTCCCCGAATTTGGGTGGA 715
325 AlaLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheG 341
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
716 AACATGTCCTCGCTCACTAGTGGCCCTGGCAGTCCAGCTCAGTTCCA 765
341 yThrThrHisIleCysGlyThrLeuIleAspAlaGlnTrpValLeuT 358
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
766 GGGTACCACCTATGTGGGGCTCCATCATCACCCCTCTGTGTGATGCTCA 815
358 hrAlaAlaHisCysPheValThrArgGluLysValLeuGluGlyTrp 374
      ||||| ||||| ||| :||| :||| :||| :||| :||| :|||
816 CTGGGCACACTGTTTATGACCTGTACCACCC.....AGTCTGTG 859
375 LysValTrpAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaLase 391
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
860 ACTGTCCAGTGGGTCTTGTGTCCCTGATGCAGACAGTCTGTGCCCTCCA 909
391 rIleAlaGlu...IleIleLeuSerAsnTyrThrAspGluAlaAspA 407
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
910 TTGTGGTGGAGAAATCATCATCACCAAGTCAACGCCAAGCGCGTGG 959
407 spTyrAspIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAla 423
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
960 GCAATGACATAGCCCTCATCAAGCTGTCCGAGCCACTCATCTTTGACGAG 1009
424 HisIleHisProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAs 440
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1010 ACCATCCAGCCCATCTGTCTGCCCACTCTGAAGAGAACTTCCCGATGG 1059
440 nGluThrCysTrpIlePheGlyPheGlyLysThrArgGluThrAspAspL 457
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1060 GAACCTGTCTGACCTCAGATGGGAGGCCACAGAGGATGGAGGTGAC. 1108
457 yThrSerProPheLeuArgGluValGlnValAsnLeuIleAspPheLys 473
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1109 ..GCCCTCCCTGCTCTGAACCCAGCGGCTGTCCCTTTAATTTCAACAAG 1156
474 LysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMe 490
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1157 ATCTGCAACACACAGGACGCTGTATGTGGTGTATCATCTCTCCCTCCATGT 1206
490 tCysAlaGlyAspLeuArgGlyArgAspSerCysGlnGlyAspSerG 507
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1207 CTGTGACGGTACTGTAGGGCGGCTGTGGACAGTGTCCAGGAGACAGTG 1256
507 lyGlyProLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyVal 523
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1257 GAGGACCCCTGGTGTGCCAGGAGGAGACTGTGGAAGTTAGTGGGTGCA 1306
524 ThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTrpH 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1307 ACAAGTTTGGAAATCGGCTGTGTGAAGTGAAGCAACAGCCTGGAGCTATAC 1356
540 rLysValThrGluValLeuProTrpIleTyrSerLysMetGluSerGluV 557
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
1357 GCGAATCACCTCTCTCTGACTGGATTTCAGAACACTTGGAGAGAGATC 1406
557 alArg 558
      :|||
1407 TGAAG 1411
seq_name: gb_pat:AX354825
seq_documentation_block:
LOCUS AX354825 2948 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0183781.
ACCESSION AX354825
VERSION AX354825.1 GI:18619556
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Meyers,R. and Macbeth,K.J.
TITLE 14094, a human trypsin family member and uses thereof
JOURNAL Patent: WO 0183781-A 1 08-NOV-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..2948
/organism="Homo sapiens"
/db_xref="taxon:9606"
628..1989
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD22993.1"
/db_xref="GI:18619557"
/translation="MGENDPPAVEAFPSRSLFGLDDLKISPVADADAAVAQILSL
PLKFFPIIVIGIIALIALAIGLGHFDCSGKRCRSFKGIELIARDGVSDCKDGE
DEYRCVRVGGONAVLOVFTAAASWKTMCSDMKWGHVAVACQALGFPVSSDNLRVSS
LEGOPREFEVSTDLHPDDKVTAHSHSVVREGCASGHVTVLQCTACGHRRCYSRIV
GNNSSLQWPRQASLOFGYHLCGGSVITPLWIIITAAHCYVDLYLPKSWTLOGLVS
LDNPAPSHLVEKIVYHSKPKRKNAAVPLINAKICNHRDVGIIISPMCLCAGYLITGG
GKVCWTSGMGATEDGDSAPVNLNAAVPLINAKICNHRDVGIIISPMCLCAGYLITGG
VDSQGDGSGPLVCQERRLWKLVGATSEFGIGCAEVNKPQVYTRVTSFLDWTHEOMERD
LKT"
BASE COUNT 696 a 759 c 732 g 738 t 23 others
ORIGIN
alignment_scores:
Quality: 744.00 Length: 561
Ratio: 2.163 Gaps: 18
Percent Similarity: 61.319 Percent Identity: 33.512
alignment_block:
US-09-879-792-12 x AX354825 ..
Align seg 1/1 to: AX354825 from: 1 to: 2948
49 ThrProProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrPr 65
||||| :||| :||| :||| :||| :||| :||| :||| :|||
364 ACTCTCAAAAAAACCAGGAATTCAGCTGTGTTAAGTCCAAGCTGAATTC 413
65 oProGlyArgAlaSerProGlyArgAlaSerProAlaGlnAlaSerProA 82
||| ||| ||||| :||| :||| :||| :||| :||| :|||
414 GGTGGG.....GGAAGGACCGGG.....C 433
82 laArgAlaSerProAlaLeuAlaSerLeuSerArgSer.....Ser 95
||||| :||| :||| :||| :||| :||| :||| :||| :|||
434 ACCGACGGCTCGGTACTTTCTGTTCTTAATTAGTTCATGCCCGTATGAG 483
96 SerGlyArgSerSerSerAlaArgSerAlaSerValThrThrSerProTh 112
||||| :||| :||| :||| :||| :||| :||| :||| :|||
484 CCAGGAAG...GGCTGTGTTTATGGGAAGCCAGTAACACTGTGGCTTAC 530
112 rArgValTyrLeuValArgAlaThrProValGly.....Alay 125
||| :||| :||| :||| :||| :||| :||| :||| :|||
```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:15:14 ; Search time 36.92 seconds
(without alignments)
589.393 Million cell updates/sec

Title: US-09-879-792-12

Perfect score: 2999

Sequence: 1 MERDSHGNSPARTPSAGAS.....TEVLPWYKMESEVFRKRS 562

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	777.5	25.9	490	1 TMS2_MOUSE	Q9jiq8 mus musculu
2	738.5	24.6	492	1 TMS2_HUMAN	O15393 homo sapien
3	731	24.4	454	1 TMS3_HUMAN	P57727 homo sapien
4	697.5	23.3	437	1 TMS4_HUMAN	Q9nrs4 homo sapien
5	664	22.1	455	1 TMS5_MOUSE	Q9er04 mus musculu
6	644.5	21.5	457	1 TMS5_HUMAN	Q9h3s3 homo sapien
7	611.5	20.4	417	1 HEP5_HUMAN	P05981 homo sapien
8	608.5	20.3	416	1 HEP5_MOUSE	O35453 mus musculu
9	592.5	19.8	416	1 HEP5_RAT	Q05511 rattus norv
10	575	19.2	1019	1 ENTK_HUMAN	P98073 homo sapien
11	574	19.1	638	1 KAL_MOUSE	P26262 mus musculu
12	572.5	19.1	638	1 KAL_HUMAN	P03952 homo sapien
13	565.5	18.9	638	1 KAL_RAT	P14272 rattus norv
14	534	18.5	1034	1 ENTK_PIG	P98074 sus scrofa
15	549	18.3	1113	1 COR1_MOUSE	Q92319 mus musculu
16	548.5	18.3	1035	1 ENTK_BOVIN	P98072 bos taurus
17	547.5	18.3	855	1 ST14_MOUSE	P56677 mus musculu
18	547	18.2	1069	1 ENTK_MOUSE	P97435 mus musculu
19	536	17.9	761	1 NEPR_MOUSE	O08762 mus musculu
20	532	17.7	625	1 FALL_HUMAN	P03951 homo sapien
21	530	17.7	1042	1 COR1_HUMAN	Q9y5g5 homo sapien
22	529	17.6	875	1 NEPR_HUMAN	P56730 homo sapien
23	520.5	17.4	422	1 DES1_HUMAN	Q9ul52 homo sapien
24	518.5	17.3	855	1 ST14_HUMAN	Q9y5v6 homo sapien
25	517.5	17.3	321	1 TRYG_HUMAN	Q9nrr2 homo sapien
26	514	17.1	812	1 PLMN_BOVIN	P06868 bos taurus
27	513.5	17.1	314	1 TEST_HUMAN	Q9y6m0 homo sapien
28	510.5	17.0	343	1 PSS8_HUMAN	Q16651 homo sapien
29	508	16.9	274	1 MCT6_RAT	P50343 rattus norv
30	506	16.9	338	1 PLMN_HORSE	P00010 equus cabal
31	505	16.8	276	1 MCT6_MOUSE	P21845 mus musculu
32	503.5	16.8	342	1 PSS8_RAT	Q9es87 rattus norv
33	503	16.8	324	1 TEST_MOUSE	Q9jhh7 mus musculu

RESULT 1
TMS2_MOUSE

ID TMS2_MOUSE STANDARD; PRT: 490 AA.
AC Q9jiq8; Q9JKC4; Q9QY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE transmembrane protein X).
GN TMPRSS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/C;
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Forvari K.S., Kellokumpu S., Kyllonen A.P., Viikho P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193;134-140(2001).
RN [2]
RP SEQUENCE FROM N.A.
RT Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC CLEAVAGE AND SECRETED (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF199362; AAF97867.1; -;
DR EMBL; AF243500; AAF64186.1; -;
DR EMBL; AF113596; AAF21308.1; -;
DR HSSP; P00761; 1AKS.
DR MGD; MGI:1354381; TmpRSS2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_recept_A.

ALIGNMENTS

34	498.5	16.6	342	1	PSS8_MOUSE
35	495	16.5	275	1	TRYT_PIG
36	495	16.5	343	1	PLMN_SHEEP
37	495	16.5	812	1	PLMN_MOUSE
38	493.5	16.5	810	1	PLMN_HUMAN
39	493	16.4	333	1	PLMN_CANFA
40	491	16.4	270	1	TRYT_MERUN
41	491	16.4	273	1	MCT7_RAT
42	488.5	16.3	273	1	MCT7_MOUSE
43	485.5	16.2	311	1	TRYG_MOUSE
44	484	16.1	273	1	TRYG_SHEEP
45	482	16.1	418	1	HATT_HUMAN

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/send an email to license@isb-sib.ch>).

CC	EMBL:	075329;	AA651784.1;	-
DR	EMBL:	AF123453;	AA371117.1;	-
DR	EMBL:	AF270487;	AAK29280.1;	-
DR	HSP:	P00763;	LDPO.	-
DR	MIM:	602060;	-	-
DR	InterPro:	IPR001314;	Chymotrypsin.	-
DR	InterPro:	IPR002172;	LDL_recept_A.	-
DR	InterPro:	IPR001190;	SRCR.	-
DR	InterPro:	IPR001254;	Trypsin.	-
DR	Pfam:	PF00057;	ldl_recept_a;	1.
DR	Pfam:	PF00089;	trypsin;	1.
DR	PRINTS:	PR00192;	CHYMOTRYPSIN.	-
DR	SMART:	SM00192;	LDLa;	1.
DR	SMART:	SM00202;	SR;	1.
DR	SMART:	SM00020;	tryp_Spc;	1.
DR	PROSITE:	PS01209;	LDLRA_1;	1.
DR	PROSITE:	PS00668;	LDLRA_2;	1.
DR	PROSITE:	PS00420;	SRCR_1;	FALSE_NEG.
DR	PROSITE:	PS0287;	SRCR_2;	1.
DR	PROSITE:	PS0240;	TRYPSIN_DOM;	1.
DR	PROSITE:	PS00134;	TRYPSIN_HIS;	1.
DR	PROSITE:	PS00135;	TRYPSIN_SER;	1.
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;			
KW	Polymorphism.			
FT	CHAIN	1	255	TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT	CHAIN	256	492	CATALYTIC CHAIN.
FT	CHAIN			TRANSMEMBRANE PROTEASE, SERINE 2,
FT	DOMAIN	1	84	CATALYTIC CHAIN.
FT	TRANSMEM	85	105	CYTOPLASMIC (POTENTIAL).
FT				SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT				(POTENTIAL).
FT	DOMAIN	106	492	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	112	149	LDL-RECEPTOR CLASS A.
FT	DOMAIN	150	242	SRCR.
FT	DOMAIN	256	492	SERINE PROTEASE.
FT	ACT_SITE	296	296	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	345	345	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	441	441	CHARGE RELAY SYSTEM.
FT	SITE	255	256	CLEAVAGE (POTENTIAL).
FT	DISULFID	113	126	BY SIMILARITY.
FT	DISULFID	120	139	BY SIMILARITY.
FT	DISULFID	133	148	BY SIMILARITY.
FT	DISULFID	172	231	BY SIMILARITY.
FT	DISULFID	185	241	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	244	365	BY SIMILARITY.
FT	DISULFID	281	297	BY SIMILARITY.
FT	DISULFID	410	426	BY SIMILARITY.
FT	DISULFID	437	465	BY SIMILARITY.
FT	CARBOHYD	213	213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	249	249	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	449	449	K -> N (IN DBSNP:1056602).
FT				/FTID=VAR_011692.
FT	MUTAGEN	255	255	R->Q: LOSS OF CLEAVAGE.
FT	MUTAGEN	441	441	S->A: LOSS OF ACTIVITY.
FT	CONFLICT	160	160	M -> V (IN REF. 3).
FT	CONFLICT	242	242	I -> L (IN REF. 1).
FT	CONFLICT	329	329	E -> Q (IN REF. 1).
FT	CONFLICT	489	489	RAD -> KAN (IN REF. 1).
SQ	SEQUENCE	492 AA;	53891 MW;	CAB44FD174A3076B CRC64;

Query Match	24.6%;	Score 738.5;	DB 1;
Best Local Similarity	31.9%;	Pred. NO. 4.9e-33;	
Matches 178;	Conservative	87;	Mismatches 194;
			Indels 99;
			Gaps 18;

QY	27	ACTPPGRA-----SPAQAASPAQAAGTTPPGRAASPAQAASPAQTTPPGRAASPGRAASPAQ 78
DB	5	SGSPAPAIQGYENHGYQYQENPYPAQTPVPT-VVEVHPAQPYPSPV-----QYAPRV 56
QY	79	ASPARPALASLSRSSGSSARSASVTTSPTRVTVLRVTPGVAVPIRSSPARSAPAT 138
DB	57	LTQA-SNPVCTQPKSPG-----TVCTSKTKKALCTLTILGTF-----94
QY	139	RATRESPGTSPLKFTWREGQKQLPLICGLVLLIALVLSLILFQFWO-----GHTGIIRY 192
DB	95	-----LVGA-----ALAAGLL-----WKFMGSKCSNSGIEC 120
QY	193	KEQRESCPKHAVRCDDVYCKLKSDGLGVCFRFDWDKSLKKIYSGSSHOWMLPTICSSWNWDS 252
DB	121	-DSSGTCINPNSWCDGVSHCPGGEDENRCVRLYGNPFIQMYSSQRKSHWHPVCQDDWNE 179
QY	253	YSEKTCQOLGFSFAHRTTEVAHRDFANFSFTLRYNST-----IQESLHRSE-CPSORYIS 306
DB	180	YGRAACRDMGYNKNFYSSQ-CIVDSSGSTSPMKLNTSAGNWDYIKKLIHSDACSSKAVYS 238
QY	307	LQCSHCGLR---AMTGRITVGGALASDKWPQVSLHFGTTHICGGTLIDAQWVLTAAHCF 363
DB	239	LRCIACGVNLSRQSRIVGSGESALPGAWPQVSLHVQNVHVCVGGSIITPEWIVTAAHCV 298
QY	364	FVTRKVKLEGG---HKVZAGTSNLHQLPEAA--STAEIIINSNTDEDDYDIALMRSLRP 418
DB	299	----EKPLNNPWHMTAFAGILRQSFMEFYGAGYQVEKVIISHPNPDSKTKNNIDIALMKLOK 354
QY	419	LTLSAHIHPACILPHWGQTFSLNFCWLTGCKTRETDDKTSPTLREVQVNLIDFKKCDNY 478
DB	335	LTFDNLVAPVCLPNPMMLOPEQLCWISGGWATEE-KGKTSEVLNAKVLIIETQRCNSR 413
QY	479	LVIDSYLTPMWCAGDLRGDRSGDGGSLPVCQEQNNRWYLAGVTSWGTGCGQRNKPVG 538
DB	414	VYVDNLITPAMICAGFLOGNVDSQGGSLPVTSKNNIWLIGDTSWGSCKAKAYRPGV 473
QY	539	YTKYTEVLPWIIYSKWESE 556
DB	474	YGNVMVFTDIYRQMR 491
RESULT 3		
TMS3_HUMAN		
ID	TMS3_HUMAN	STANDARD; PRT; 454 AA.
AC	P57727;	
DC	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	01-MAR-2002 (Rel. 41, Last annotation update)	
DE	Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease	
DE	TAGC-12) (Tumor associated differentially-expressed gene-12 protein).	
GN	TPRSS3 OR TAGD12 OR ECHOS1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
ON	[1]	
RP	SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).	
RC	TISSUE=Ovarian carcinoma;	
RX	MEDLINE=20521358; PubMed=11068177;	
RA	Underwood L.J., Thigmona K., Tanimoto H., Beard J.B., Schneider E.N.	
RA	Wang Y., Parmley T.H., O'Brien T.J.;	
RT	"Ovarian tumor cells express a novel multi-domain cell surface serine	
RT	protease.";	
RL	Biochim. Biophys. Acta 1502:337-350(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D), AND VARIANT ILE-53.	
RX	MEDLINE=20578749; PubMed=11137999;	
RA	Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrast R.	
RA	Gulponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,	
RA	Younis F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,	
RA	Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Famir B.,	
RA	Antonarakis S.E.;	
RT	Insertion of beta-satellite repeats identifies a transmembrane	


```
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "Molecular cloning of mouse type 4 spinesin.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1, 2, 3 AND 4 (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB016229; BAB20276.1; -
DR EMBL; AB016230; BAB20277.1; -
DR EMBL; AB016423; BAB20278.1; -
DR EMBL; AB041037; BAB40328.1; -
DR MGD; MGI:1933407; Tmprs5.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001190; SRCR.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 49
FT TRANSMEM 50 70
FT SIGNAL-ANCHOR (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT SRCR.
FT SERINE PROTEASE.
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CLEAVAGE (POTENTIAL).
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT MISSING (IN ISOFORM 2).
FT MISSING (IN ISOFORM 3).
FT MISSING (IN ISOFORM 1).
FT MISSING (IN ISOFORM 1).
FT GGLVFEAWKP -> MEAQVGLLWV (IN ISOFORM 1).
FT D -> G (IN REF. 1; BAB20277).
FT SEQUENCE 455 AA; 49632 MW; 5CFC31789C6899AA CRC64;
Query Match 22.1%; Score 664; DB 1; Length 455;
Best Local Similarity 31.4%; Pred. No. 4.4e-29;
Matches 144; Conservative 79; Mismatches 174; Indels 62; Gaps 13;
```

```
QY 139 RATRESPGTSLPKFTWREGQKQLPLI-----GCVLLLI-----ALVSVLLILFQ 182
DB 14 RCTEEGAGPGI--FRMELGDORQISOSORWCCLQRCVILGVLLLAGAGIASWLLIVLY 71
QY 183 FWOGHT-----GIRYKEQRESCPKHAVRCD-----GVVDCKLKASDELGCYRFWDKS 229
DB 72 LWPAAFPSISGTLQEEEMTLNCP--GVSCBEEELLPSLPKTVSPFRINGEDL----- 119
QY 230 LKIIYSGSHQWLPIGSSNNWDSYSEKTCQQLGF--ESAHRTEVA-----HRDFANSF 281
DB 120 LLOVQVRRPDLVLLVCHEGSPALGHMICKSLGHIRLTQHKAVNLSLDIKLNRSQEAFLS 179
QY 282 SILRYNSTIOESLHRS-ECPSORYISLQCHSCHGLRAWTRIGVGALASDSKWPQVSLHF 340
DB 180 A--RPGGLVEEAKWPSANGPSGRIVSLKSCGAPRLASIRVGGQAVASGVPWQASVML 237
QY 341 GTTHICGGTLIDAQWLTAACFFVTRKVLGKWKVA-----GTSNLHOLPEAASTAEI 395
DB 238 GSRHTCGASVLAHPHVVVTAACHMYSFRLSKLSNRVHAGLVSHGAVRQHQ---GTWVEKI 294
QY 396 IINSNTYDEEDDDYDIALMRSLKPLTSLAHIPACLPMHGQTFSLNETCWTGFGKTRTD 455
DB 295 IPHPLYSAQNHDYDVALQLRTPINESDTVDVCLPAKEQYFPWGSQCWYSGWGHDPDSH 354
QY 456 DKTSPEFLREVQVNLIDFKKNDLYVDVSYLTPRMCAGDLRGDRDSCQSGGPLYCEQN 515
DB 355 THSSDTLQDTMVPVPLSTHLCSNCSMTSGALTHRLMCAGYLDGRADACQDGGGLVPCSG 414
QY 516 NRWYLAGVTSWGTGCGRNKPGVYTKVTEVLPWIYSKME 554
DB 415 DTWHLVGVVSWGRGCAEPNRPGVYAKVAEFLDWIHTVQ 453
RESULT 6
TMSS5_HUMAN STANDARD; PRT; 457 AA.
AC Q9H3S3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 5 (BC 3.4.21.-) (Spinesin).
GN TMRPS5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX PubMed=11741986;
RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
RT human spinal cord.;
RL J. Biol. Chem. 276:0-0(2001).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
CC neurons, in their axons, and at the synapses of motoneurons in the
CC spinal cord.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB028140; BAB20375.1; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001190; SRCR.
```

DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS02087; SRCR_2; FALSE_NEG.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein. 1 49
FT DOMAIN 50 70 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 71 457 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 112 207 SRCR.
FT DOMAIN 218 457 SERINE PROTEASE.
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).
FT SITE 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT DISULFID 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;

Query Match 21.5%; Score 644.5; DB 1; Length 457;
Best Local Similarity 31.2%; Pred. No. 4.8e-28;
Matches 148; Conservative 69; Mismatches 18; Indels 69; Gaps 12;

Qy 126 PIRSPASAPATRATRESPGTSPLKFT----WRGQKPLIGVLLIALVWS--LII 179
Db 10 PHEQYAEFGPGIFRAEPGDQHPISQAVCWRSWRGCAVLGALLAGVGSWLLV 69
Qy 180 LLFQWQGHGTIRYKQRESCPKHVRCDGVVCKLKSDE--IGC----- 221
Db 70 LY-----LCPAASQPISGT-----LDDEITLSCSEASAEALLPALPKT 109
Qy 222 --VREDWKSILKIYSGSHQWLPICSSWNDSYSEKTCQQLGFSAHRTTEVAHRDFAN 279
Db 110 VSFRIINSEDFLEAQRDPRWLLVCHGWSPALGLQICWSLGH-----LRLTHKGYN 163
Qy 280 SFSILRYNST-----IQESLH-RSECPQRYISLQCSHCLRAMTGRIVGGA 325
Db 164 -LTDIKLNSQEEFAQLSPRLGGFLEEAQPNRNCTSGQVSLRCSECGARPLASRVGGQ 222
Qy 326 LASDSKWPQVSLFHGTTHICGGTLLDAQWVLTAAHCFVTRKYLEGKVVYAG-TSNLH 384
Db 223 SVAPGRWPQASVALGFRHTCGGSLAPRWVVTAAHCHSFRLARLSSWRVHAGLVSHA 282
Qy 385 QLPEAASTAE-IINSNYTDDEDDYDIALMRSLKPLTSLAHHPACLPWHGQTFSLNETC 443
Db 283 VRPHQGLVERIIPPLYSQAQNHVDVALLRLQTALNFSDTVGAYCLPAKEQHFKGSRG 342
Qy 444 WITGFGKTRTDDKTSPLRELVQVNLIDFKKCNDFLVVDSYLTTPRMCAAGDLRGGRDSCQ 503
Db 343 WYSGHGHTPHSHVTSDDMLQDVTVPVLFSTQLCNSSCVTSGALTPRMCAGYLDGRADACQ 402
Qy 504 GDSGGPLCEONNRWYLAGVTSWGTGCGQRNKPQYVTVKTEVLPWIKSMSEV 557
Db 403 GDSGGPLVCPDGTWRLVGVVSWGRACAEPNHPGVYAKVAEFLDWHDTAQDSL 456

RESULT 7
HEPS_HUMAN
ID HEPS_HUMAN STANDARD; PRT; 417 AA.
AC P05981;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine 1).
DE 1).
GN HPN OR TMPRSS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88209431; PubMed=2835076;
RA Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RT "A novel trypsin-like serine protease (hepsin) with a putative
RT transmembrane domain expressed by human liver and hepatoma cells.";
RL Biochemistry 27:1067-1074(1988).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=91358502; PubMed=1885621;
RA Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA Chou S.H., Kurachi K.;
RT "Hepsin, a cell membrane-associated protease. Characterization,
RT tissue distribution, and gene localization.";
RL J. Biol. Chem. 266:16948-16953(1991).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=93348237; PubMed=8346233;
RA Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
RT "Hepsin, a putative cell-surface serine protease, is required for
RT mammalian cell growth.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
CC -!- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST
CC LEVEL IN LIVER.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
EMBL; M18930; AAA36013.1; -;
EMBL; X07732; CAA30558.1; -;
EMBL; X07002; CAA30058.1; -;
PIR; S00845; S00845.
HSP; P00734; 2HNT.
MEROPS; S01.224; -;
MIM; 142440; -;
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR001190; SRCR.
InterPro; IPR001254; Trypsin.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00202; SR; 1.
SMART; SM00202; Tryp_SPC; 1.
PROSITE; PS02040; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 162 NON-CATALYTIC CHAIN (POTENTIAL).
FT CHAIN 163 417 CATALYTIC CHAIN (POTENTIAL).

```
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 18 44 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 45 417 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 163 417 SERINE PROTEASE.
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 153 277 INTERCHAIN (BY SIMILARITY).
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 322 338 BY SIMILARITY.
FT DISULFID 349 381 BY SIMILARITY.
FT CARBOHYD 112 112 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 417 AA; 45011 MW; B2086FF661E551D7 CRC64;

Query Match 20.4%; Score 611.5; DB 1; Length 417;
Best Local Similarity 36.0%; Pred. No. 2.6e-26;
Matches 134; Conservative 62; Mismatches 147; Indels 29; Gaps 9;

QY 209 VVCKLKSD--LGCVRFDWDKSLKIIYSGSSHOWLPICSSNNWDSYSEKTCQQLGFESA 266
D 39 IVALLLRSDOEPLYPQVSSADARLWFDKTEGTWRLCCSSRNARVAGLSCEMGFLRA 98
QY 267 HRTTEVAHRD-----FANSFSLRYNSTIQESLHRSCEPQRYISLQCHSGGLRAM 317
D 99 LTHSELVDRTAGANGTSGFFCVDEGRPLHTQRLLEVISVCDPCRGRLAAICQDCGRRKL 158
QY 318 -TGRIVGGALADSKWPQVSLHFGTHTCGGTLIDAQWLVTAACHFFVTRKVLGKVV 376
D 159 PVDRIVGGRTSLGRWPQVSLRYDGAHLCCGSLSGDWLVTAACH-FPERNKLVSRRV 217
QY 377 YAGT----SNLHOLPEAAIAEIIINSY-----TDEDDYDIALMRLSKPTLSAHIHP 427
D 218 FAGAVAQASPHGL--QLGVQAVVHYGYLPFRDPNPEENSNDIALVHLSPLPLEYIQP 275
QY 428 ACLPMHQFSLNETCWTGTFGKTRTDKTSPLREVQVNLIDFKKNDYLVYDYLTP 487
D 276 VCLPAAGALVDGKICTVTVGWTGNTQYGOQAG-VLQEARVPIISNDVNCAGDFYGNQIKP 334
QY 488 RMCAGDLRGGRDSCGDSGGPLVCE----QNNRWYLAGVTSWGTGCGQRNPGVYTKVT 543
D 335 KMFACGYPEGIDACGDSGGFVCEDSISRTPRWRLCGIVSWGTCALAQKPGVYTKVS 394

QY 544 EYVLPWYISKMES 555
D 395 DFEWIFQAICKT 406

RESULT 8
HEPS_MOUSE STANDARD; PRT; 416 AA.
AC O35453;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-).
GN HPN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98058912; PubMed=9395459;
RA Vu T.-K.H., Liu R.W., Haakma C., Tomasek J.J., Howard E.W.;
RT "Identification and cloning of the membrane-associated serine
protease, hepsin, from mouse preimplantation embryos.";
RL J. Biol. Chem. 272:31315-31320(1997).
CC -!- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
OF CELL MORPHOLOGY.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AF030065; AAB84221.1; -.
DR HSSP; P00734; 2HNT.
DR MEROPS; S01.224; -.
DR MGD; MGI:1196620; Hpn.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001190; SRCR.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
FT CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 162 416 SERINE PROTEASE.
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 132 276 INTERCHAIN (BY SIMILARITY).
FT DISULFID 187 203 BY SIMILARITY.
FT DISULFID 321 337 BY SIMILARITY.
FT DISULFID 348 380 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 416 AA; 44739 MW; 432194FF4004F848 CRC64;

Query Match 20.3%; Score 608.5; DB 1; Length 416;
Best Local Similarity 36.8%; Pred. No. 3.7e-26;
Matches 140; Conservative 60; Mismatches 135; Indels 45; Gaps 12;

QY 209 VVCKLKSD--LGCVRFDWDKSLKIIYSGSSHOWLPICSSNNWDSYSEKTCQQLGFESA 266
D 38 IVTILLQSDOEPLYPQVSSADARLWFDKTEGTWRLCCSSRNARVAGLSCEMGFLRA 97
QY 267 HRTTEVAHRD-----ANSFS-----ILRYNSTIQESLHRSCEPQRYISLQCHSGC 312
D 98 -----LAHSELVDRTAGANGTSGFFCVDEGGLPLAQRLDVISVCDPCRGRLTATCQDC 152
QY 313 GLRAM-TGRIVGGALADSKWPQVSLHFGTHTCGGTLIDAQWLVTAACHFFVTRKVL 371
D 153 GRKLPVDRIVGGQDSSLGRWPQVSLRYDGHLCGSLSGDWLVTAACH-FPERNRLV 211
QY 372 EGWKVYAG---TSNLHQLPEAA--SIAEIIINSY-----TDEDDYDIALMRLSKPL 419
D 212 SRWRVFAAGAVARTS-----PHAVQGVQAVIYHGGYLPFRDPTIDENSNDIALVHLS 266
QY 420 TLSAHIHPACLPWHGQTFSLNETCWTGTFGKTRTDKTSPLREVQVNLIDFKKNDYL 479
D 267 PLTEYIQVCLPAAGQALVDGKVTCTVGTGNT-QFVGQQAQVQLQEARVPIISNEVCNSPD 325
QY 480 VYDSYLTTPRMCAAGDLRGGRDSCGDSGGPLVCEQN-----NRWYLAGVTSWGTGCGQRNK 535
D 326 FYGNQIKPKMFCAGYPEGGIDACGDSGGFVCEDSISGTSRWRLCGIVSWGTCALARK 385
QY 536 PGVYTKVTEVLPWYISKMES 555
```



```
FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 134 134 Q -> E (IN REF. 3).
FT CONFLICT 732 732 S -> P (IN REF. 3).
FT CONFLICT 754 771 SQCLQSLRLRLOCNHKS -> RRNAKNEIDALSPIILIA
FT SEQUENCE 1019 AA; 112923 MW; B6AAA245F6DA563 CRC64;

Query Match 19.2%; Score 575; DB 1; Length 1019;
Best Local Similarity 33.3%; Pred. No. 5.3e-24;
Matches 133; Conservative 56; Mismatches 156; Indels 52; Gaps 12;

QY 187 HTGI-----RYKEQRESCPKHVRCDGVVDDCKLSDGLGCVRFWDKSLKLYSGSSH 239
DB 636 HLGIPCKADHFQCKNGECVPLVNLCDGHLHCEGDSDEADCVRF-----ENGTN 686

QY 240 Q-----WLPICSSWNWDSYSEKTCQOLGFEFAHRTTEVAHRDFANSFILRYNS 288
DB 687 NGLVRFRIOSIHTACAEWNNTQISNDVCQLLGLSGNSKPIFTSD---GGPFVKLNT 743

QY 289 TIQESL---HRSECPQRSYISLQCSH---CGLRA---MTGRIVGGALASDSKWPQVWSLH 339
DB 744 APDGHLLTFSQCLQSLRLRLOCNHKSCGKKLAAQDITPKIVGGSNAKEGANWVVGLY 803

QY 340 FGTHICGGTLIDAOVLTAHCFVTRREKVGKVVYAGTSNHLQLPERAS-----I 392
DB 804 YGRLLCGASLVSSDLVSAHCVY-GRNLEPSKWTALIG---LHMSNLTSPQTVPRLI 859

QY 393 AEIILNSNYDEEDDYDIALMRLSKPTLSAHIHPACLPKHGQTFSENETCWITGFKTR 452
DB 860 DEIVINPHYRRRRKNDNDIAMHLEFKVNYTDYIQICLPEENQVFPFGRNCISAGWG-TV 918

QY 453 ETDDKTSPEFREVOVNLIDFKKNDYLVDSYLPTRMWCAGDLRGDRSCGDSGGPLVC 512
DB 919 VYOGTTANILQEADVPILSNRCCQQMP-EYNITENMICAGYEEGGIDSCQDGGGPLMC 977

QY 513 EQNNRWLAGVTSMTGCGGQNRKPGVYTKVTEVLPWIYS 551
DB 978 QENNRWFLAGVTSFGYKCALPNRPGVYARVSRFTWLIQS 1016

RESULT 11
KAL_MOUSE STANDARD; PRT; 638 AA.
AC P26262;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)
DE (Kininogenin) (Fletcher factor).
GN KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RP STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=91090844; PubMed=2264928;
RA Saidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.;
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
```

```
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M58588; AAA63393.1; -.
CC PIR; A36557; KOMSPL.
CC HSP; P20231; IAAO.
CC MEROPS; S01_212; -.
CC MGD; MGI:102849; Kik3.
CC InterPro; IPR000177; Apple.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR003014; PAN.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00024; PAN; 4.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00005; APPLIEDOMAIN.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00223; APPLE; 4.
CC SMART; SM00020; TRYPSIN; 1.
CC PROSITE; PS00495; APPLE; 4.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
CC Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
CC Repeat.
CC SIGNAL.
FT CHAIN 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT DOMAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 O-LINKED (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
CC SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;
```

Query Match	19.1%;	Score 574;	DB 1;	Length 638;
Best Local Similarity	33.7%;	Pred. No. 3.9e-24;		
Matches 131;	Conservative 61;	Mismatches 95;	Indels 102;	Gaps 15;
QY	196	RESCPKHVRCDGVV-----DCKLSDELGC---VRFDWDKSLKTKYSGSSHQWLPICS	246	
DB	319	QETCTK-TIRCQFIYSLLPQDCK---EEGCKCSLRLSDGSPTRITYG-----	363	
QY	247	SNWDSYSBKTCQOLGPFESAHRTTEVAHRDFANSFSLRYNSTIQESLARHSECPQRSYIS	306	
DB	364	MGSSGYSRLKCLV--DSPDCTKI-----	387	
QY	307	LQCHSGCLRAMTGRIVGALASDKSPWQVSLH---FGTHICGGTLIDAQWVLTAAHCF	363	
DB	388	-----NARIVGNTASLGEWPMQVSLQVQLVSOHLCCGGSTIGRQWVLTAAHCF	436	
QY	364	FVTRKVLG-----WKVYAGTSLNHLPE---AASIAETIIINSYTOEDDDYDIALMR	414	
DB	437	-----DGIYPDVWRIYGGILSLSEIKETPSSRIKELIHQEVKVGSEGYDIALIK	488	
QY	415	LSKPLTLSAHIHPACLPMHGOTFSLNCTWITGFKTRETDDKTSPLREVQVNLIDFKK	474	
DB	489	LQPLNLTETQKPICLPSKADNTIYTNCTWGTGTYKE-QGETQNTLQKATIPLVPEE	547	
QY	475	C-----NDYLVYDVSILTPRMKAGDLRGDRSCQDGGPLVCEQNNRWYLAGVTSWGTGC	530	
DB	548	CQKRYRDIYI-----NKQMCAGYKGGTACKDGGSGPLVCKHSGRWQLVGTISWGECC	602	
QY	531	GQRNPGVYTKVTEVLPWYISKME-SEVR 558		
DB	603	GRKDPGVYTKVSEYMDWILEKTSQSSDVR 631		
RESULT 12				
ID	KAL_HUMAN	STANDARD;	PRT;	638 AA.
AC	P03952;			
DT	23-OCT-1986	(Rel. 02, Created)		
DT	23-OCT-1986	(Rel. 02, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)			
DE	(Kininogenin) (Fletcher factor).			
GN	KLK3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86243359; PubMed=3521732;			
RA	Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;			
RT	"Human plasma prekallikrein, a zymogen to a serine protease that			
RT	contains four tandem repeats.";			
RL	Biochemistry 25:2410-2417(1986).			
RN	[2]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=91152016; PubMed=1998666;			
RA	McMullen B.A., Fujikawa K., Davie E.W.;			
RT	"Location of the disulfide bonds in human plasma prekallikrein: the			
RT	presence of four novel apple domains in the amino-terminal portion of			
RT	the molecule.";			
RL	Biochemistry 30:2050-2056(1991).			
CC	-1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT			
CC	ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING			
CC	TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM			
CC	HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN			
CC	SYSTEM BY CONVERTING PRORENIN INTO RENIN.			
CC	-1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES			
CC	THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,			
CC	AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE			
CC	CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.			
CC	-1- DISEASE: DEFECTS IN KLK3 ARE THE CAUSE OF FLETCHER FACTOR			
CC	DEFICIENCY, A BLOOD COAGULATION DEFECT.			
-1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.				
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.				
This SWISS-PROT entry is copyright. It is produced through a collaboration				
between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
the European Bioinformatics Institute. There are no restrictions on its				
use by non-profit institutions as long as its content is in no way				
modified and this statement is not removed. Usage by and for commercial				
entities requires a license agreement (see http://www.isb-sib.ch/announce/				
or send an email to license@isb-sib.ch).				
EMBL; M13143; AAA60153.1; -				
PIR; A00921; KOHUP.				
PIR; A37939; A37939.				
HSSP; P00766; ICHG.				
MEROPS; S01.212; -				
MIM; 229000; -				
InterPro; IPR000177; Apple.				
InterPro; IPR001314; Chymotrypsin.				
InterPro; IPR003014; PAN.				
InterPro; IPR001254; Trypsin.				
Pfam; PF00024; PAN; 4.				
Pfam; PF00089; trypsin; 1.				
PRINTS; PR00005; APPLEDOMAIN.				
PRINTS; PR00722; CHYMOTRYPSIN.				
SMART; SM00223; APPLE; 4.				
SMART; SM00020; TRYP-SPC; 1.				
PROSITE; PS00495; APPLE; 4.				
PROSITE; PS0240; TRYPSIN_DOM; 1.				
PROSITE; PS00134; TRYPSIN_HIS; 1.				
PROSITE; PS00135; TRYPSIN_SER; 1.				
KW	Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;			
KW	Fibrinolysis; Blood coagulation; Inflammatory response; Liver;			
KW	Repeat.			
FT	SIGNAL	1	19	
FT	CHAIN	20	390	
FT	CHAIN	391	638	
FT	DOMAIN	20	105	
FT	DOMAIN	110	195	
FT	DOMAIN	200	285	
FT	DOMAIN	291	376	
FT	DOMAIN	389	621	
FT	CARBOHYD	127	127	
FT	CARBOHYD	308	308	
FT	CARBOHYD	396	396	
FT	CARBOHYD	453	453	
FT	CARBOHYD	494	494	
FT	ACT_SITE	434	434	
FT	ACT_SITE	483	483	
FT	ACT_SITE	578	578	
FT	DISULFID	21	104	
FT	DISULFID	47	77	
FT	DISULFID	51	57	
FT	DISULFID	111	194	
FT	DISULFID	137	166	
FT	DISULFID	141	147	
FT	DISULFID	201	284	
FT	DISULFID	227	256	
FT	DISULFID	231	237	
FT	DISULFID	292	375	
FT	DISULFID	318	347	
FT	DISULFID	322	328	
FT	DISULFID	340	345	
FT	DISULFID	383	503	
FT	DISULFID	419	435	
FT	DISULFID	517	584	
FT	DISULFID	548	563	
FT	DISULFID	574	602	
FT	DISULFID	638	AA;	
FT	DISULFID	71369	MM;	
FT	DISULFID	E62F9C1053838FB4		
FT	DISULFID	SEQUENCE		
QY	638	AA;	71369	MM;
QY	E62F9C1053838FB4			
Query Match		19.1%;	Score 572.5;	DB 1; Length 638;

	Best Local Similarity	34.5%	Pred. No.	4.7e-24;	
	Matches	132; Conservative	53; Mismatches	101; Indels	97; Gaps
Qy	196	RESCPKHVRCDGVV-----DCKLKSDELGC-VRFDMDKSLTKT---YSGSSHOWLPIC	245		
	:	: : :	: :	: :	: :
Dz	319	QETCTK-MIRCOFFYSLLPEDC--KEEKCKFLRLSMDGSPTRAYTCQGSSGYSLRLC	375		
Qy	246	SSNWNDSYEKTQQOGLGFESAHRTVEAHRDPANSFSILRYNSTIQESLHRSECPQRSYI	305		
	:	: : :			
Dz	376	NTGDNSVCVTTKT-----	387		
Qy	306	SLOCSHCGLRAMTGRIYVGALASDKWPQVLSLHFGTT---HICGGTLIDAQWYLTAABC	362		
	:	: : :	: :	: :	: :
Dz	388	-----STRIVGGTNSSGEWPMQVSLQVKLTQAQRHLCCGSGLIGHOWLYLTAABC	435		
Qy	363	FVTVREKVLEG-----WKVVAGTSNLHQLPE--AASTAEITINSNYTDEDDYDIALLM	413		
	:	: : :	: :	: :	: :
Dz	436	F-----DGLPLQDVWRISGIINLSLDITKDPPFQSQIKELIIHQNVKVEGNHDIALI	487		
Qy	414	RLSKPLTSAHHIPACLPMHGOTTSNLTCWNITGFQKTRETDDKTSPLPREVQVNLI DFK	473		
	:	: : :	: :	: :	: :
Dz	488	KLOAPLINTFEQKPICLPISKGTSTIYNVWVTGWGFSKERGE-IQNLIKVNPIPLVTNE	546		
Qy	474	KCNDYLVYDSV-LNPRMMCAGDLGGRRDCOGDGSGGPLVCEONNRWYLAGVTSWGTCGGQ	532		
	:	: : :	: :	: :	: :
Dz	547	EQQR--YQDYKKITQRMVCAGYKEGGKDACKDGGGPLVCRKHNGMWRLVGITSWGEGCAR	604		
Qy	533	RNKPGVYTKTEVLPTWIYSKMES	555		
	:	: : :	: :	: :	: :
Dz	605	REQPGVYTKVAEYMDWILEKTQS	627		
RESULT	13				
KAL_RAT		STANDARD;	PRT:	638 AA.	
ID	KAL_RAT				
AC	PI4272;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last annotation update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)				
DE	(kininogenin) (Fletcher factor).				
GN	PK.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
NCBI_TaxID	10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE-91129236; PubMed-1993180;				
RA	Beaubien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,				
RA	Seldah N.G.;				
RT	"Gene structure and chromosomal localization of plasma kallikrein.";				
RL	Biochemistry 30:1628-1635(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RP	MEDLINE-90091743; PubMed-2598771;				
RA	Seldah N.G., Ladenheim R., Mbikay M., Hamelin J., Lutfalla G.,				
RA	Rougeon F., Lazure C., Chretien M.;				
RT	"The cDNA structure of rat plasma kallikrein.";				
RL	DNA 8:563-574(1989).				
CC	-!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT				
CC	ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING				
CC	TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM				
CC	HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN				
CC	SYSTEM BY CONVERTING PRORENIN INTO RENIN.				
CC	SUBUNIT: THE ZYMEOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES				
CC	THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,				
CC	AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE				
CC	CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.				
CC	-!- SIMILARITY: CONTAINS 4 APPLE DOMAINS.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.				
CC	-----				

FT	DISULFID	206	225	BY SIMILARITY.	16-OCT-2001 (Rel. 40, Last sequence update)	DT
FT	DISULFID	219	236	BY SIMILARITY.	16-OCT-2001 (Rel. 40, Last annotation update)	DT
FT	DISULFID	658	670	BY SIMILARITY.	Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-converting enzyme) (Corin) (Low density lipoprotein receptor related protein 4).	DE
FT	DISULFID	667	692	BY SIMILARITY.	CRN OR LRP4.	DE
FT	DISULFID	787	911	INTERCHAIN (BY SIMILARITY).	Mus musculus (Mouse).	GN
FT	DISULFID	825	841	BY SIMILARITY.	OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	OS
FT	DISULFID	925	992	BY SIMILARITY.	NCBI_TaxID=10090;	OC
FT	DISULFID	956	971	BY SIMILARITY.	[1]	OX
FT	DISULFID	982	1010	BY SIMILARITY.	SEQUENCE FROM N.A.	RN
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).	MEDLINE=98429596; PubMed=9756624;	RP
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .) (POTENTIAL).	Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;	RX
FT	CARBOHYD	170	170	N-LINKED (GLCNAC. . .) (POTENTIAL).	"A novel low-density lipoprotein receptor-related protein with type II membrane protein-like structure is abundant in heart.";	RA
FT	CARBOHYD	194	194	N-LINKED (GLCNAC. . .) (POTENTIAL).	J. Biochem. 124:784-789(1998).	RT
FT	CARBOHYD	283	343	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).	CC
FT	CARBOHYD	343	350	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- SUBCELLULAR LOCATION: Type II membrane protein.	CC
FT	CARBOHYD	403	403	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART.	CC
FT	CARBOHYD	455	455	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.	CC
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.	CC
FT	CARBOHYD	518	518	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- SIMILARITY: CONTAINS 2 FRIZZLED (FZ) DOMAINS.	CC
FT	CARBOHYD	549	549	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.	CC
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	697	697	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	701	701	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	721	721	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	740	740	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	804	804	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	863	863	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CC
SQ	SEQUENCE	1034 AA; 114776 MW; 0388C64CF64CC368 CRC64;			CC	CC
Query Match					EMBL; AB013874; BAA34371.1; -	DR
Best Local Similarity					HSP; P01130; IAJJ.	DR
Matches 136; Conservative 56; Mismatches 155; Indels 52; Gaps 14;					MGD; MGI:1349451; Lrp4.	DR
QY	187	HTGI-----RYKEQRESCPKHVRGCVVDDCKLSDGLGCVRFWDKSLKTIYSGSSH	239		InterPro; IPR001314; Chymotrypsin.	DR
Db	651	HLGPECKEDNFCENGECVLLVNLCDGFSHCKDGSDEACVRF-----LNGTAN	701		InterPro; IPR000024; Fz.domain.	DR
QY	240	Q-----WLPICSSNNDSYSEKTCQQLGFSAHRTTEVAHRDFANSFS--ILRY	286		InterPro; IPR002172; LDL_recept_A.	DR
Db	702	NSGLVQPRIQSIWHTACAEENWTQTSDDVCQLGLGTGN-----SSMPFSSGGGPFVKL	756		InterPro; IPR001190; SRCR.	DR
QY	287	NSTIQESL---HRSECPQRSYISLQCSH--CG----LRAMTGRIVGGALASDSKWPQVFS	337		InterPro; IPR001254; trypsin.	DR
Db	757	NTAPNGSLILTASQCPEDSLILQCNHSCGKKQVAQEVSPKIVGGNDSREGAWPWVA	816		Pfam; PF01392; Fz; 2.	DR
QY	338	LHFGTTHICGTLIDAQWLTAHCFVTRKVLGKVVYAG---TSNLHOLPEAAS--I	392		Pfam; PF00057; ldl_recept_a; 7.	DR
Db	817	LYNGQLLCASLSVRDLVSAACHVY-GRNLEPSKWKAILGLHMTSNLTS-POIVTRLI	874		Pfam; PF00089; trypsin; 1.	DR
QY	393	AEIINSNYTDEDDYDIALMRSLKPIITLSAHIHPACLPMHGQTFSLNETCWTIFGKTR	452		PRINTS; PR00722; CHYMOTRYPSIN.	DR
Db	875	DEIVINPHYRRRRKSDSIAMHLEFKYNTDYIQPICLPEENQVFPFGICSIAGWGVK-	933		PRINTS; PR00261; LDLRECEPTOR.	DR
QY	453	ETDDKTSFPLREVQVNLIDFKCNDYLDVDSYLTPTMWCAGDLRGGRDSCGDSGGPLVC	512		SMART; SM00063; FRI; 2.	DR
Db	934	IYOGSPADILQAEADVPLLSNEKCCQQMP-EYNITENMMWCAEGEGIDSCQDGGGLPMC	992		SMART; SM00192; LDLa; 7.	DR
QY	513	EONRWYLAGVTSWGTGCGQRNKPQVTKVTEVLPVWYS	551		SMART; SM00202; SR; 1.	DR
Db	993	LENNRWLLAGVTSFGYOCALPNRPGVAVRPKFTWITQS	1031		SMART; SM00020; Tryp_spc; 1.	DR
RESULT 15					PROSITE; PS50038; Fz; 2.	DR
CORI_MOUSE					PROSITE; PS01209; LDLRA_1; 6.	DR
ID	CORI_MOUSE	STANDARD;	PRT; 1113 AA.		PROSITE; PS50068; LDLRA_2; 7.	DR
AC	Q92319;				PROSITE; PS00420; SRCR_1; FALSE_NEG.	DR
DT	16-OCT-2001 (Rel. 40, Created)				PROSITE; PS50287; SRCR_2; 1.	DR
					PROSITE; PS00240; TRYPsin_DOM; 1.	DR
					PROSITE; PS00134; TRYPsin_HIS; FALSE_NEG.	DR
					PROSITE; PS00135; TRYPsin_SER; 1.	DR
					Hydrolase; Serine protease; Transmembrane; Signal-anchor;	KW
					Glycoprotein; Repeat.	KW
					DOMAIN 1 112	FT
					TRANSMEM 113 133	FT
					CYTOPLASMIC (POTENTIAL).	FT
					SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN (POTENTIAL)).	FT
					EXTRACELLULAR (POTENTIAL).	FT
					FZ 1.	FT
					LDL-RECEPTOR CLASS A 1.	FT
					LDL-RECEPTOR CLASS A 2.	FT
					LDL-RECEPTOR CLASS A 3.	FT
					LDL-RECEPTOR CLASS A 4.	FT
					FZ 2.	FT
					LDL-RECEPTOR CLASS A 5.	FT

FT	DOMAIN	683	721	LDL-RECEPTOR CLASS A 6.
FT	DOMAIN	722	757	LDL-RECEPTOR CLASS A 7.
FT	DOMAIN	758	853	SRCL
FT	DOMAIN	869	1113	SERINE PROTEASE.
FT	ACT_SITE	910	910	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	959	959	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1052	1052	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID	857	979	BY SIMILARITY.
FT	DISULFID	895	911	BY SIMILARITY.
FT	DISULFID	1022	1037	BY SIMILARITY.
FT	DISULFID	1048	1077	BY SIMILARITY.
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	202	202	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	208	208	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	298	298	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	317	317	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	411	411	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	481	481	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	519	519	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	537	537	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	635	635	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	719	719	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	765	765	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	828	828	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	970	970	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1113 AA;	122984 MW;	B845B2C5F20DD8EC CRC64;
Query Match 18.3%; Score 549; DB 1; Length 1113;				
Best Local Similarity 35.9%; Pred. No. 1.4e-22;				
Matches 137; Conservative 60; Mismatches 135; Indels 50; Gaps 18;				
Qy	199	CPRHVRCDGVDCKLKSDELGCYRFDWD---	KSLKIIY-SGSSHOWLPICSSNWNDYS	254
Db	735	CVPRDLWCDGWDGSDSDDEWGCVTLSKNGNSSLLTVHKSAKEHH---	VCADGWRETL	791
Qy	255	EKTCQQLGFESAHRTTEVAHRD-----	FANSFSLRYNSTIOESL-HRSECP	305
Db	792	QLACKQMGLEPSVTKLIPGQEGQWLRLYPNENL---	NGSTLOELLVYRHSCPSRSEIS	849
Qy	307	LQCS--HCGLR---AMTGRIVGGALASDKWPQVSLHFGTT-HICGGTLIDAQWLTAA	360	
Db	850	LLCSKQDCGRRPAARMNKRIILGRTSRPRGWPQC	SLQSEPSGHICGCVLI	909
Qy	361	HCFFVTREKVLG-----WKVYAGTSLNHQ---	LPEAASTAEIINSNTDEDDYDIA	411
Db	910	HCFF-----EGREDADVWKVVFGINNLDHPSGFMQTRFVKTLILHPRYSRAVVYD	DIS	961
Qy	412	LMRLSKPLTLSAHIHPACLPMHGQTFSLNETCWTGFKTRTDKTS	SPF-LREVOVNL	470
Db	962	VVELSDDINETSIVRVPCLSPPEYLEPDIYCYITGWHM---	GKMPFKLQGEVRI	1017
Qy	471	DFKCNLYLYD-SYLTPRMACADLRGSDSGGGLVCEQ-NNRWYLAGVTSWGT	528	
Db	1018	PLEQCQSY--FDMKTIITNRMICAGYESGTVDSCMGDSGGPLVCERPGQWTLFGLT	SWGS	1075
Qy	529	CGQOR-NKPGVYTKVTEVLP	PHI	549
Db	1076	VCFSKVLGPGVYNSVYFVGWI	1097	

Search completed: August 20, 2002, 04:22:55
Job time: 461 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:41:39 ; Search time 52.41 Seconds
(without alignments)
1030.380 Million cell updates/sec

Title: US-09-879-792-12
Perfect score: 2999
Sequence: 1 MERDSHGNSPARTPSAGAS.....TEVLPWTYKMESEVRFRKS 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	611.5	20.4	417	1 S00845	hepsin (EC 3.4.21.
2	592.5	19.8	416	1 S33777	hepsin (EC 3.4.21.
3	575	19.2	1019	1 A56318	enteropeptidase [E
4	574	19.1	638	1 KQHUP	plasma kallikrein
5	572.5	19.1	638	1 KQHUP	plasma kallikrein
6	568.5	19.0	855	2 JC7731	membrane-bound arg
7	568.5	19.0	855	2 JC7775	membrane type-seri
8	565.5	18.9	638	1 KORTPL	plasma kallikrein
9	554	18.5	1034	1 A53663	enteropeptidase [E
10	549	18.3	1113	2 JF0315	low-density lipopr
11	548.5	18.3	1035	1 A43090	enteropeptidase [E
12	536	17.9	761	2 JC5759	brain-specific ser
13	532	17.7	625	1 KPHU1	coagulation factor
14	514	17.1	812	1 PLB0	plasmin (EC 3.4.21
15	510.5	17.0	343	1 A57014	proctasin (EC 3.4.
16	508	16.9	274	2 JC4171	tryptase (EC 3.4.2
17	506.5	16.9	455	2 A61545	plasmin (EC 3.4.21
18	505	16.8	276	2 A38654	mast cell proteina
19	495	16.5	460	2 B61545	plasmin (EC 3.4.21
20	495	16.5	812	1 PLMS	plasmin (EC 3.4.21
21	493.5	16.5	810	1 PLHU	plasmin (EC 3.4.21
22	491	16.4	270	2 S56160	mast cell tryptase
23	488.5	16.3	273	2 A47246	tryptase (EC 3.4.2
24	480	16.0	275	2 C35863	tryptase (EC 3.4.2
25	480	16.0	786	1 A47547	serine proteinase
26	480	16.0	790	1 PLPG	plasmin (EC 3.4.21
27	479	16.0	275	2 A35863	tryptase (EC 3.4.2
28	479	16.0	415	1 A34170	acrosin (EC 3.4.21
29	478	15.9	275	2 B35863	tryptase (EC 3.4.2

30	477.5	15.9	418	2 A37344	acrosin (EC 3.4.21
31	477	15.9	275	2 A32410	tryptase (EC 3.4.2
32	473	15.8	437	2 S18407	acrosin (EC 3.4.21
33	472.5	15.8	810	2 B30848	plasmin (EC 3.4.21
34	471.5	15.7	4548	1 S00657	apoptein(a) (EC
35	470	15.7	436	2 JX0172	acrosin (EC 3.4.21
36	468.5	15.6	1524	2 T30337	apolipoprotein(a)
37	466.5	15.6	1420	2 A32869	apolipoprotein(a)
38	466	15.5	421	2 S29599	acrosin (EC 3.4.21
39	466	15.5	655	1 A46688	hepatocyte growth
40	457.5	15.3	274	2 A45754	tryptase (EC 3.4.2
41	457	15.2	431	2 S47538	acrosin (EC 3.4.21
42	456	15.2	810	2 I46260	plasmin (EC 3.4.21
43	453.5	15.1	421	1 S11674	acrosin (EC 3.4.21
44	447	14.9	434	1 A35005	u-plasminogen acti
45	438	14.6	237	2 S68702	tryptase (EC 3.4.2

ALIGNMENTS

RESULTS

1

S00845

hepsin (EC 3.4.21.) - human

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999

C:Accession: S00845

R:Levtus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.

Biochemistry 27, 1067-1074, 1988

A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane

A:Reference number: S00845; MUID:88209431

A:Accession: S00845

A:Molecule type: mRNA

A:Residues: 1-417 <EX>

A:Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064

C:Genetics:

A:Gene: GDB:HNP; TMRSS1; hepsin

A:Cross-references: GDB:I35685; OMIM:142440

A:Map position: 19q11-19q13.2

C:Superfamily: hepsin; trypsin homology

C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein

F:23-45/Domain: transmembrane #status predicted <TMN>

F:163-400/Domain: trypsin homology <TRY>

F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted

F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 20.4%; Score 611.5; DB 1; Length 417;

Best Local Similarity 36.0%; Pred. No. 2.6e-26;

Matches 134; Conservative 62; Mismatches 147; Indels 29; Gaps 9;

QY 209 VVDCIKLSDP--LGCVRFWDKSLKIIYSGSSHOWLPICSSNNWDSYSEKTCQOLGFESA 266

DB 39 IAVLURSDQEPYLPVQVSSADARLMVFQKTEGTWLLCSSRNARVAGLSCCEMFLRA 98

QY 267 HRTTEVAHRD-----FANSFSLRYNSTIQESLHRSCEPQSYRIISLQCCHCGLRAM 317

DB 99 LTHSELDTAGANGTSGFECVDEGRPLPHTQRLLEVISVDCPCPRGLAICODCCGRKL 158

QY 318 -TGRIYGGALASDKWPQVSLHFGTTHICGGFLIDAQWVLTAAHCFVTRKVLGKWKV 376

DB 159 PVDRIYVGGRTSLGRWPQVSLRYDGAHLGSGLLSGDWLTAAHC-FPERNRVLSRWV 217

QY 377 YAGT---SNLHQLPEAASTAEIISNY-----TDEDDYDIALMRLSKPLTSLSHIHP 427

DB 218 FAGAAQAQAPHLG--OLGQVAVYHGGYLPFRDPNSENNDIALVHLSPLPLTYIQP 275

QY 428 ACLPMHGQTFSLNETCWTGFGKTRTDKTSFPLREVOVNLIDFKKNDLYVDSYLRP 487

DB 276 VCLPAAQALVDGKICTVTGNGNTQYGGQAG-VLGEARVPIISNDVCGADFYGNQIKP 334

QY 488 RMMACGLRGRDSCQDGGPLVCF---QNNRWYLAGVTSWGTGCGQRNKPQVYTKVT 543

Db 335 KMFCAGYPEGIDACQDGGPFVCEDSISRTPRRLCGIVSWGTCALAAQKPGVYTKVS 394

Qy 544 EVLPWIKSMES 555
: ||: :::

Db 395 DFREWIFOAIKT 406

RESULT 2

S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R:Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A>Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A:Reference number: S33777; MUID:93305733
A:Accession: S33777
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <FAR>
A:Cross-references: EMBL:X70900; NID:957928; PIDN:CAA50256.1; PID:957929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TMN>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match 19.8%; Score 592.5; DB 1; Length 416;
Best Local Similarity 36.3%; Pred. No. 2.8e-25;
Matches 138; Conservative 59; Mismatches 138; Indels 45; Gaps 12;

Qy 209 VVDCKLKSDE--LCGVREDNDKSLLLKIYSGSSHOWLPICSSNNWSDSYSEKTCQQLGFESA 266
:||:||:| | :| | :| | :| | | | :| | :| | :| | :| | :| |
Db 38 IVTILLRSDDPLEYQVQSPOGSRULLVDKTEGTWRLLCCSSRNARVAGIGCEEMLFLRA 97
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Qy 267 HRTEVAHRDP-----ANSFS-----ILRYNSTIOESLRHSECPSORYISLQCSHC 312
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 98 ----LAHSELDRVRTAGANGTSGFVCVDEGLPLAQRLLDVISVCDCPRGRELTATQCQC 152
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Qy 313 GLRAM-TGRIVGGALASDKPWPQVSLHFHTTHICGGTLIDAQWLTAACHFPVTREKVL 371
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 153 GRKLPDVRIVGGDSSILGRWPWPQVSLRDYDGHLCGGSLLSGDWLVLTAAHC-FPERNRVL 211
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Qy 372 EGWKVIAG----TSLNLHOLPEAA--SIAEIIINSNY-----TDEEDDYDIALMRLSKPL 419
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 212 SRWRVFAGAVARTS-----PHAVQLGVQAIVHYGGYLPRDPTIDENSNDIALVHLSSSL 266
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Qy 420 TLSHHIIPACLPMHGQTSLNETCWITFGKTRETDDKTSPLREVQVNWLIDFKKCNXYL 479
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 267 PLTEYIOPCVLPAAGQALVDGKVCTVTWGNT-QFYGOQAVVLQEAREVPPIISEVNCSPD 325
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Qy 480 VYDSLPLPRMWCADGLRGDRSCDGSGPLVCEQ-----NNRWYLAGVTSWTGTCGQRNK 535
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 326 FYGNQIRPKMFCAGYPEGGIDACQDGGSHFVCEDSISRRLCGIVSWGTCALARK 385
:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Qy 536 PGVYTKVTEVLPWIYSKMES 555
||||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 386 PGVYTKVIDFEWIFOAIAKT 405

RESULT 3

A56318
enteropeptidase (EC 3.4.21.9) precursor - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C>Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
C:Accession: A56318; B43090
R:Kitamoto, Y.; Veille, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A>Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic

A:Reference number: A56318; MUID:95234679

A:Accession: A56318

A:Molecule type: mRNA

A:Residues: 1-1019 <K12>

A:Cross-references: GB:U09860; NID:g746412; PIDN:AAC50138.1; PID:g746413

R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease co

A:Reference number: A43090; MUID:94329561

A:Accession: B43090

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 749-1019 <K12>

A:Cross-references: GB:U09860

C:Comment: The mechanism of association with the membrane of the intestinal brush bor

otated below) or with amino-terminal myristoylation of the heavy chain.

C:Genetics:

A:Gene: GDB:PRSS7

A:Cross-references: GDB:384083; OMIM:226200

A:Map position: 21q21-21q21

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig

ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv

ducts.

C:Function:

A:Description: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Clis repeat homology: LDL receptor ligand-binding

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymoge

F:1-784/Product: enteropeptidase heavy chain #status predicted <HC>

F:22-38/Domain: transmembrane #status predicted <TM>

F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:342-504/Domain: MAM homology <MAM>

F:526-631/Domain: Clr/Clis repeat homology <Clr>

F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:785-1019/Product: enteropeptidase light chain #status predicted <LC>

F:785-1014/Domain: trypsin homology <TR>

F:785-1014/Domain: trypsin homology <TR>

F:116-147/179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding sit

F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted

F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match	19.2%	Score 575;	DB 1;	Length 1019;
Best Local Similarity	33.3%	Pred. No. 6.2e-24;		
Matches 133;	Conservative 58;	Mismatches 156;	Indels 52;	Gaps 12;

QY	187	HTGI-----RYKEQRESCPKHAVRCGDYVVDCKLSDLGCVRFMDKSLLLKIYSGSSH	239
DB	636	HLGIPECKADHFQCKNGECVPLVNLCDGHLHCEGDSDEADCVRF-----FNGTNN	686
QY	240	Q-----WLPICSSWNWDSYSEKTCQQLGFESAHTTEVAHRDFANSFSLRYSN	288
DB	687	NGLVRFRIQSIWHTACAEWNTTQISNDVCQLLGLSGNSKSPIFSTD---GGPFVKLNT	743
QY	289	TIQESL---HRSECPQRYISLQCSH---CGLRA---MTGRIVGGALASDPKWPQVSLH	339
DB	744	APDGHLLTSPQQLQSDLSRLQCNHKSCKGLAAQDITPKIVGGSNAKGAEPWVGLY	803
QY	340	FGTHICGGTLIDAQWYLTAAHCFEFTVREKVLGKWKYAGTSNLHQLPEAAS-----I	392
DB	804	YGGRLCCGALSIVSDWLVSAAHCY-GRNLEPSKWTAILG---LHMKNLITSPQTPRLI	859
QY	393	AEIINSNYTDEEDDYDIALMRLSKPLTLSAHITHPACLPMHGQTFSLNEWCWITGCKTR	452
DB	860	DEIVINPHYNRKNDNDIAMHLEFFKNVYTDYTOPICLPEENQVFPFGRNCSTAGWG-TV	918
QY	453	ETDDKTSPFLEQVNLIDFKKNDYLVYDSYLTPTRMCAAGDLRGDRSDCGSGGFLVC	512
DB	919	YVOGTTANILQEADVPLLSNERCOQMP-EYNTENNMCAGYEEGGIDSCQDGGGFLMC	977
QY	513	EQNRWYLAGVTSNGTCGGGORNKPGVYTKYTVLVPWYIS	551
DB	978	QENRWYLAGVTSFGYKCALPNRPGYVARYSRTEWIQS	1016

C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four
A:Reference number: A00921; MUID:86243359
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHU>
A:Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence
A:Reference number: A37939; MUID:91152016
A:Accession: A37939
A:Molecule type: protein
A:Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;1
'260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,
525;538-551;562,'X',584-587;573,'X',575-576;578-583,'X',585;592-604 <MCX>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciproc
inogen and may also play a role in the renin-angiotensin system by converting proreni
C:Genetics:
A:Gene: GDB:KLK3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-4q35
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-638/Product: plasma kallikrein #status predicted <MAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,
F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimenta
F:318-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 19.1%; Score 572.5; DB 1; Length 638;
Best Local Similarity 34.5%; Pred. No. 5.2e-24;
Matches 132; Conservative 53; Mismatches 101; Indels 97; Gaps 13;

QY 196 RESCPKHAHVRCGVV-----DCKLKSDELGC--VRFDWDKSLKLI---YSGSSHQWLPIC 245
DB 319 QETCTK-MIRCOFFTYSLLPEDC--KECKCKFLRLSMDGSPTRIAVGTQSGSYSLRLC 375
QY 246 SNWSDSYSEKTCQQLGFSAHRTTEVAHRDFANSFILRYNSTIOESLHRCPCSQRYI 305
DB 376 NTGDSNVCVTKT-----
QY 306 SLQCSHCLRAMTGRIVGGALASDSKWPQVSLHFGTT---HICGTLIDAQWVLTAAHC 362
DB 388 -----STRIVGTNSGWPQVSLQVLTQRLHCLGSLGHQWVLTAAHC 435
QY 363 FVTRKVLLEG-----WKVYAGTSLNHLPE---AASIAEIIINSNYTDEDDYDIALM 413
DB 436 F-----DGLPLQDVWRIYSGILNSDITKTPPSQIKEIIHHQYKVSSEGNHDI 487
QY 414 RLKSPLTLSAHIHPACPLMHGQTFSLNETCWITGFGKTRTDKTPFPLREVQVNLIDFK 473
DB 488 KLQAPLNYTEFQKPICLPSKADNTIYTCWVTGWTGVSKEKGE-IQNILQKNVPIVLTNE 546
QY 474 KNDYLVYDSY-LTPRMCAAGDLRGGRDSCGDSGGLVCEQNNRWYLAGVTSWGTGCG 532
DB 547 ECKR--YQDYKITQRMVCAVYKGGKDGKDGSGGLVCEQNNRWYLAGVTSWGTGCG 604

RESULT 4
KOMSP
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A36557
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compariso
A:Reference number: A36557; MUID:91090844
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: GB:M58588; NID:g200358; PIDN:AAA63393.1; PID:g200359
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 19.1%; Score 574; DB 1; Length 638;
Best Local Similarity 33.7%; Pred. No. 4.3e-24;
Matches 131; Conservative 61; Mismatches 95; Indels 102; Gaps 15;

QY 196 RESCPKHAHVRCGVV-----DCKLKSDELGC--VRFDWDKSLKLIYSGSSHQWLPICS 246
DB 319 QETCTK-TIRCOFFTYSLLPQDCK---EEGCKSLRLSDGSPTRIVG----- 363
QY 247 SNWSDSYSEKTCQQLGFSAHRTTEVAHRDFANSFILRYNSTIOESLHRCPCSQRYIS 306
DB 364 MQSGSYSLRLCKLV--DSPDCTKI----- 387
QY 307 LQCSHCLRAMTGRIVGGALASDSKWPQVSLH---FGTHICGTLIDAQWVLTAAHCF 363
DB 388 -----NARIVGTNSAGLSPWPQVSLQVLTQSLHCLGSLIIGRWVLTAAHCF 436
QY 364 FVTRKVLLEG-----WKVYAGTSLNHLPE---AASIAEIIINSNYTDEDDYDIALMR 414
DB 437 -----DGLPYDVWRIYGGILSELTKTPPSRIKELIIHQYKVSSEGNYDIALIK 488
QY 415 LSKPILTLSAHIHPACPLMHGQTFSLNETCWITGFGKTRTDKTPFPLREVQVNLIDFK 474
DB 489 LQPLNYTEFQKPICLPSKADNTIYTCWVTGWTGVTRE-QGETNQLQKATIPLPNNEE 547
QY 475 C-----NDYLVYDSY-LTPRMCAAGDLRGGRDSCGDSGGLVCEQNNRWYLAGVTSWGTGC 530
DB 548 COKKYRDYVI-----NKQMICAGYEGGTACKGDSGGLVCEQNNRWYLAGVTSWGTGC 602
QY 531 GORNPQGVYTKVTEPLPIYSKME--SEVR 558
DB 603 GRKDPQGVYTKVSEYMDWILEXTQSDVR 631

RESULT 5
KOHUP
plasma kallikrein (EC 3.4.21.34) precursor - human
N:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)

A:Reference number: JE0315; MUID:98429596

A:Accession: JE0315

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1113 <M>

A:Cross-references: DBJ:AB013874; NID:93869144; PID:BAA34371.1; PID:93869145

C:Superfamily: trypsin homology; LDL receptor ligand-binding repeat homology

F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>

F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>

F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>

F:869-1087/Domain: trypsin homology <TRY>

Query Match 18.3%; Score 549; DB 2; Length 1113;

Best Local Similarity 35.9%; Pred. No. 1.7e-22;

Matches 137; Conservative 60; Mismatches 135; Indels 50; Gaps 18;

QY 199 CPKHAVRCGVVDCIKLDELGCVRPDMW---KSLIKIY-SGSSHOWLPICSSNNWDSYS 254

DB 735 CVPRDLWCDGWDCSDSDGCVTLKNGSSLLTVHKSKEHH---VCADGWRETLS 791

QY 255 ETKCQOLGFESAHRTTEVAHRD-----FANSFSILRYNSTIQESL-HRSECPQRYIS 306

DB 792 OLACKOMGLGEPSTVKLIPGQEQWLRYPNWNEL--NGSTIQELLYVRHSCPSRSEIS 849

QY 307 LOCS--HCLGR---AWTGRIVGALASDSKWPQVSLHFGTT-HICGGTLIDAQWVLTAA 360

DB 850 LUCSKODGRRPAARNKRIIGRTSRPRGWPQCQLQSEPSGHICGCVLIAKKVVLTVA 909

QY 361 HCFVTRERKVLGEG-----WKVYAGTSNLHQ---LPEAASIAEIIINSNTDEEDYDIA 411

DB 910 HCF-----EGREDADWKVVGFLNNLDHPGFMQTRFVKTILLHPYSRVVDYDIS 961

QY 412 LMRSLKPLTSAHIHPACLMHGQTFSLNETCWITGFGKTRTDKTSFP-LREVQVNL 470

DB 962 VVELSDINDIETSVYRVCPLSPPEVLEPDYCYITGWGHM---GNKMPFKLQGEVRII 1017

QY 471 DFKKNDYLVD--SYLTPRMKGADLRGRDSCQDGGGLPLVCEQ--NNWYLAGVTSWT 528

DB 1018 PLEQCOQSY--FMKTTNRMCIAVESGTVDSQMGDSGGPLVCERPQGQWTLFGLTWSGS 1075

QY 529 GCGQR--NKPGVYTKVTEVLPWI 549

DB 1076 VCFSKVLGPGVSNVSYFYGWI 1097

RESULT 11

A43090 enteropeptidase (EC 3.4.21.9) precursor - bovine

N:Alternate names: enterokinase

C:Species: Bos primigenius taurus (cattle)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A43090; A48874; A61436

R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo

A:Reference number: A43090; MUID:94329561

A:Accession: A43090

A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1035 <KIT>

A:Cross-references: GB:U09859; NID:9746410; PID:AA840026.1; PID:9746411

R:Experimental source: Small intestine

R:LaVallie, E.R.; Rehemtulla, A.; Racic, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.;

J. Biol. Chem. 268, 23311-23317, 1993

A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of b

A:Reference number: A48874; MUID:94043122

A:Accession: A48874

A:Molecule type: mRNA

A:Residues: 801-1035 <LAV>

A:Cross-references: GB:L19663; NID:9416131; PID:AAAL6035.1; PID:9416132

A>Note: parts of this sequence, including the amino end of the mature protein, were c

R:Light, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A:Reference number: A61436; MUID:92189715

A:Accession: A61436

A:Molecule type: protein

A:Residues: 801-807, 'Y', 809-827 <LIG>

C:Comment: The mechanism of association with the membrane of the intestinal brush bor

embrane attachment using a signal-anchor sequence.

C:Comment: Conversion from membrane-bound to soluble forms may involve further proces

C:Complex: mature enteropeptidase is variously reported to contain two (heavy and lig

hlike-linked

C:Function:

A:Description: cleaves propeptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Clis repeat homology; LDL receptor ligand-binding

C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane prot

F:22-38/Domain: transmembrane #status predicted <TM>

F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>

F:198-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:358-520/Domain: MAM homology <MAM>

F:542-647/Domain: Clr/Clis repeat homology <CLR>

F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>

F:801-1030/Domain: trypsin homology <TRY>

F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bin

F:788-912,826-842,926-993;957-972,983-1011/Disulfide bonds: #status predicted

F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 18.3%; Score 548.5; DB 1; Length 1035;

Best Local Similarity 34.1%; Pred. No. 1.7e-22;

Matches 128; Conservative 54; Mismatches 144; Indels 49; Gaps 12;

QY 206 CDGVVDCIKLSDFELGCVRFDMKSLIKYSGSSHQ-----LNGITDSSGLVQFRIQSIHWVACAEWTTQIS 254

DB 678 CDGFFHCKDGSDEAHCVR-----LNGITDSSGLVQFRIQSIHWVACAEWTTQIS 728

QY 255 ETKCQOLGFESAHRTTEVAHRDFANSFS-----ILRYNSTIQESL---HRSECPQRYIS 306

DB 729 DDVQQLLGLGTGNSSVP-----TFSTGGPYVNLNAPNGSLILTPSQOCLDSLIL 780

QY 307 LOCSH--CGLRAMT---GRIVGALASDSKWPQVSLHFGTHICGGTLIDAQWVLTAA 360

DB 781 LQCNKSKCGKILVTQEVSPKIVGGSDSREGAMPWVVALYFDQOQVCGASLSVRDVLVSA 840

QY 361 HCFVTRERKVLGKWKVYAG---TSNLHQLP-EAASIAEIIINSNTDEEDDYDIALMRLS 416

DB 841 HCIV-GRNPEPSKWKAVLGLHMASNLTSPOIETRLDIOIVINPHYKRRKNNDIAMHLE 899

QY 417 KPLTISAHIHPACLMHGQTFSLNETCWITGFGKTRTDKTSFPFLREVQVNLIDFKKN 476

DB 900 MKVNVTDYIQCIPLENQVFPFGRICSIAGWG-ALIVQGSTADVLQEADVPLLSNEKQP 958

QY 477 DYLVVDSYLTTPRMKGADLRGRDSCQDGGGLVCEQNNRWYLAGVTSWGTCCGQNRKP 536

DB 959 QOMP-EYNTENNVCAGYEAGGVDSQCGSGGLPMQCENNRWLLAGVTSFGYQCALPNRP 1017

QY 537 GVTYKTEVLPVLI 551

DB 1018 GYARVPRETEWIIQS 1032

RESULT 12

JC5759

brain-specific serine proteinase (EC 3.4.21.-) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 31-Mar-2000

Db 255 CLKTSGLPSTRIKSKALSQFSLOSCHRSIPVFCSSFYHDTDFLGEELDIVAAKSH 314
QY 256 KTCQQLGFSAHRTTEVAHRDFANSFILIYNSTIQESLHRSCEPQSRQYISLQCSHGCLR 315
Db 315 EACQKL-----CTNAVRCQF---FTYTPAASCNEG--KGKC-----YKLSSNGSPTK 358
QY 316 AMTG-----RIVGALASDSKWPQVSLHFGTT-----HIC 346
Db 359 ILHGRGIGSYTLRLCKMNECTTKIKPRIVGTGASVSGEWPQVTLH--TTSPTQRHLIC 416
QY 347 GGPLDAQWLVTAHCHFF-VTREKVLGKWKVYAGTSLNHLQHPAAS---IAETIIINSNT 402
Db 417 GGSIIQNQWILTAACHFCYGVESPIL--RVYSGLNQSEIKEDTFFGVEIILHDQYK 473
QY 403 DEEDDYDIALMRUSKPLTUSAHITPACPLMHGQTFSLNETCWITGFKGTRETDKTSFPL 462
Db 474 MABSGYDIALLKLETTVNTDSORPCLPSKGRNVIYDCWWTGNG-VYRKLRDKTQNTL 532
QY 463 REVQVNLIDFKKNDVLYVDSY-LTPRMCAGLDGRDSCGDSGGLVCEQNNRWYLA 521
Db 533 QKAKIPLVNTEECQKR--YRGHKITHKMICAGYREGGDKACKGDSGGLPSCKRHVWHLV 590
QY 522 GVTSWGTGCGQNRKPGVYTKVTEVLPWIYSKMES 555
Db 591 GITSWEGCAQREPGVYTVNVEYVDWILEKTOA 624
RESULT 14
PLBO
plasmin (EC 3.4.21.7) precursor - bovine
N:Alternate names: plasminogen
C:Species: Bos primigenius taurus (cattle)
C:Date: 30-Sep-1987 #sequence_revision 28-Apr-1995 #text_change 18-Jun-1999
C:Accession: S45046; A25835; I45961; S03736
R:Bergrund, L.; Andersen, M.D.; Petersen, T.E.
submitted to the EMBL Data Library, May 1994
A:Description: Cloning and characterization of the bovine plasminogen cDNA.
A:Reference number: S45046
A:Accession: S45046
A:Molecule type: mRNA
A:Residues: 1-812

A:Cross-references: EMBL:X79402; NID:9494962; PIDN:CAA55939.1; PID:9494963
A:Experimental source: liver
A:Note: It is uncertain whether Met-1 or Met-8 is the initiator
R:Schaller, J.; Moser, P.W.; Danneberg-Muller, G.A.K.; Rosselet, S.J.; Kampf, U.; Rick
Eur. J. Biochem. 149, 267-278, 1985
A:Title: Complete amino acid sequence of bovine plasminogen. Comparison with human plasmin
A:Reference number: A25835; MUID:8520331
A:Accession: A25835
A:Molecule type: protein
A:Residues: 27-334,'D',336-515,'H',517-554,'L',556-812 <SCH>
R:Malinowski, D.P.; Sadler, J.E.; Davie, E.W.
Biochemistry 23, 4243-4250, 1984
A:Title: Characterization of a complementary deoxyribonucleic acid coding for human and
A:Reference number: I45961; MUID:8502331
A:Accession: I45961
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 706-743,'R',745-812 <MAL>
A:Cross-references: GB:K02935; NID:9163551; PIDN:AAA30714.1; PID:9163552
R:Brumsholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M.;
Eur. J. Biochem. 114, 465-470, 1981
A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of human,
A:Reference number: S03735; MUID:81212097
A:Accession: S03736
A:Molecule type: protein
A:Residues: 27-83 <BRU>
C:Function:
A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
ns the walls of the graafian follicle; also activates the urokinase-type plasminogen act
A:Pathway: fibrinolysis
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology
C:Keywords: duplication; fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma;

F:1-26/Domain: signal sequence #status predicted <SIG>
F:8-103/Domain: plasminogen-related protein precursor homology <PLPH>
F:27-812/Product: plasminogen #status experimental <PRO>
F:27-103/Domain: activation peptide #status experimental <APT>
F:104-583,584-812/Product: plasmin #status experimental <MAT>
F:104-583/Domain: plasmin chain A #status experimental <ACH>
F:110-188/Domain: kringle homology <KR1>
F:192-269/Domain: kringle homology <KR2>
F:282-359/Domain: kringle homology <KR3>
F:384-461/Domain: kringle homology <KR4>
F:485-564/Domain: kringle homology <KR5>
F:584-812/Domain: plasmin chain B #status experimental <BCH>
F:584-805/Domain: trypsin homology <TRY>
F:56-80,60-88,110-188,131-171,159-183,192-269,195-323,213-252,241-264,282-359,303-342
bonds: #status predicted
F:315/Binding site: carboxydrate (Asn) (covalent) #status experimental
F:365/Binding site: carboxydrate (Ser) (covalent) #status experimental
F:624,667,763/Active site: His, Asp, Ser #status predicted
Query Match 17.1%; Score 514; DB 1; Length 812;
Best Local Similarity 35.5%; Pred. No. 9.9e-21;
Matches 122; Conservative 42; Mismatches 126; Indels 54; Gaps 9;
QY 245 CSSNWDNSYSEK-----TCQOLGFESAHRTEVAHRDFANSFSLRYNSSTIQESLHR 296
Db 485 CMIGTKSYRGKATTVAGVPCOEAAQBPHQ-----HSITPETNPQGLER 532
QY 297 SEC-----PSQRYISLQCS---HCGI-----RAMTGRIVGGALASD 329
Db 533 NYCNRPDGDVNGPWCYTMNPKPFYDCVPQCBSSFDCKGPKVEPKKSGRIVGGCVSKP 592
QY 330 SKWPQVSVLHFGTHTCGGTLDAQWLVTAHCHFFVTRKVLGKWKVYAGTSLNHLQLEA 389
Db 593 HSWPQVSVLRRSRHFCGGTGLSPKWLTAACH--LDNIALSFFKVLGAHN--EKVRE 648
QY 390 ASAEIIINSNTDEDDYDIALMRUSKPLTUSAHITPACPLMHGQTFSLNETCWITGFG 449
Db 649 QSVQEIPIV-SRIFREPSQADIALLKLSREAITITKVIPACLPPTNPYVVAARTECIYTWG 707
QY 450 KYRETDKTSFPLEVQVNLIDFKKNDVLYVDSYLTTPRMCAGLDGRDSCGDSGGP 509
Db 708 ETQGTGE--GLLKEALPVIENKVCNRYLDGRVKPTCLCAGHLIGTDCSQDSGGP 765
QY 510 LVCEQNNRWYLAGVTSWGTGCGQNRKPGVYTKVTEVLPWIYSKM 553
Db 766 LVCFEKDKVILQGTSMGLGCARPKNKPGVYVSPYVPWPIETM 809
RESULT 15
A57014
proctasin (EC 3.4.21.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 18-Jun-1999
C:Accession: A57014; A54866
R:Yu, J.X.; Chao, L.; Chao, J.
J. Biol. Chem. 270, 13483-13489, 1995
A:Title: Molecular cloning, tissue-specific expression, and cellular localization of
A:Reference number: A57014; MUID:95286644
A:Accession: A57014
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-343 <RES>
A:Cross-references: GB:L41351; NID:g862304; PIDN:AAC41759.1; PID:g862305
A:Experimental source: prostate
A:Note: parts of this sequence were determined by protein sequencing
R:Yu, J.X.; Chao, L.; Chao, J.
J. Biol. Chem. 269, 18843-18848, 1994
A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purificatio
A:Reference number: A54866; MUID:94308140
A:Accession: A54866
A:Molecule type: protein
A:Residues: 45-64 <YUA>

Search completed: August 20, 2002, 04:16:24
Job time: 2085 sec

CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
CC systemic sclerosis and aberrant wound repairs, circulatory disorders
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
CC dyostolity, sclerodactyly and teangiectasis, dermatological disorders
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
CC caused by ocular neovascular disease, corneal graft neovascularisation,
CC macular degeneration, retinopathy of prematurity, retrolental
CC fibroplasia and corneal neovascularisation. The nucleic acids of the
CC invention are also used in gene therapy. The invention also provides
CC method for screening compounds that modulate angiogenesis.
XX
SQ Sequence 1689 BP; 367 A; 529 C; 470 G; 323 T; 0 other;

alignment_scores:
Quality: 2991.00 Length: 562
Ratio: 5.332 Gaps: 0
Percent Similarity: 99.822 Percent Identity: 99.822
alignment_block:
US-09-879-792-12 x AAD05796 ..
Align seg 1/1 to: AAD05796 from: 1 to: 1689

1 MetGluArgAspSerHisGlyAsnAlaSerProAlaArgThrProSerAl 17
1 ATGGAGAGGGACACCCACGGGAATGCATCTCCACGAGAAGACACCTTCAGC 50
17 aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProProGlyArgA 34
51 TGGAGCATCTCCAGCCAGGCATCTCCAGCTGGGACACCTCCAGCGCGG 100
34 laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
101 CATCTCAGCCAGGCATCTCCAGCCAGGCATCTCCAGCTGGGACACCT 150
51 ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProProGl 67
151 CCGGGCCGGGCATCTCCAGCCAGGCATCTCCAGCTGGGACACCTCCAGC 200
67 yArgAlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaArgA 84
201 CCGGGCATCTCCAGCCGGGCATCTCCAGCCAGGCATCTCCAGCCCGG 250
84 laSerProAlaLeuAlaSerLeuSerArgSerSerGlyArgSerSer 100
251 CATCTCCGGCTCTGGCATCACTTCCAGGTCCCTCATCCGGCAGGTCA 300
101 SerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeuVa 117
301 TCCGCCAGGTGAGCTCGGTGACAACTCCCAACACAGAGTGTACCTTGT 350
117 lArgAlaThrProValGlyAlaValProIleArgSerSerProAlaArgS 134
351 TAGAGCAACACCACTAGTGGGGCTGTACCCATCCGATCATCTCCGCCAG 400
134 erAlaProAlaThrArgAlaThrArgGluSerProGlyThrSerLeuPro 150
401 CAGCACCAGCAACAGGGCCACCCAGGGAGAGCCAGGTACGAGCCCTGCC 450
151 LysPheThrTrpArgGluGlyGlnLysGlnLeuProLeuIleGlyCysVa 167
451 AAGTTCACCTGGCGGGAGGGCCAGAACGACGCTACCGCTCATCGGTGCGT 500
167 lLeuLeuLeuIleAlaLeuValValSerLeuLeuIleIlePheGlnPheT 184
501 GCTCCCTCATTCGCTGGTGGTTCGCTCATCATCTCTTCAGTTCT 550
184 rpGlnGlyHisThrGlyIleArgTyrLysGluGlnArgGluSerCysPro 200

551 GGCAGGGCCACACAGGGATCAGGTACAGGAGCAGAGGGAGAGAGCTGTCCC 600
201 LysHisAlaValArgCysAspGlyValValAspCysLysLysSerAs 217
601 AAGCACGCTGTTCCCTGTGACGGGGTGGTGGACTGCAAGCTGAAGAGTGA 650
217 pGluLeuGlyCysValArgPheAspTrpAspLysSerLeuLeuLysIleT 234
651 CGAGCTGGCTGCTGAGGTGTGACTGGGACAGTCTCTGCTTAAATCT 700
234 yrSerGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsn 250
701 ACTCTGGGTCTCCATCAGTGGCTTCCATCTGTAGCAGCACTGGAAT 750
251 AspSerTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaH 267
751 GACTCTACTCAGAGAAGACCTGCCAGCAGCTGGGTTCGAGAGTGTCTCA 800
267 sArgThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleL 284
801 CCGGACAACCGAGGTGCCACAGGGATTTGCCAACAGCTTCTCAATCT 850
284 euArgTyrAsnSerThrIleGlnGluSerLeuHisArgSerGluCysPro 300
851 TGAGATACAACTCCACCATCCAGGAAGCCCTCCACAGGTCTGAATGCCCT 900
301 SerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMe 317
901 TCCAGCGGTATATCTCCCTCCAGTGTTCACACTGCGGACTGAGGGCAT 950
317 tThrGlyArgIleValGlyGlyAlaLeuAlaSerAspSerIlystrPrpT 334
951 GACCGGCGGATCTGGGAGGGGCGCTGGCTCGGATAGCAAGTGGCCTT 1000
334 rpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeu 350
1001 GGCAAGTGAGTCTGCACCTCGGCACACCCACACATCTGTGGAGGCACGCTC 1050
351 lIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhePheValThrAr 367
1051 ATTGACGCCAGTGGGTGCTCACTCGCGCCGCTCTCTTCGTGACCCG 1100
367 gGluLysValLeuGluGlyTrpLysValTyrAlaGlyThrSerAsnLeuH 384
1101 GGAGAAGTCTCTGAGGGCTGGAAGGTGTACGGGGCCACCAGCAACCTGC 1150
384 isGlnLeuProGluAlaAlaSerIleAlaGluIleIleAsnSerAsn 400
1151 ACCAGTTGCTGAGGCAGCCTCCATTGCCGAGATCATCAACAGCAAT 1200
401 TyrThrAspGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLy 417
1201 TACACCGATGAGGAGGAGCTATGACATCGCCCTCATGGGCTGTGCAA 1250
417 sProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProMethisG 434
1251 GCCCTGACCCCTGTCCGCTCACATCCACCTGCTTGCTCCCCATGCATG 1300
434 lyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLys 450
1301 GACAGACCTTTAGCTCAATGAGACCTGCTGGATCAGAGCTTTGGCAAG 1350
451 ThrArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnVa 467
1351 ACCAGGGAGACAGATGACAAGACTCCCTTCTCCGGGAGGTGACAGGT 1400
467 lAsnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrAspSerT 484
1401 CATATCATCGACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGTT 1450
484 yrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAsp 500
1451 ACCTTACCCCAAGGATGATGCTGCTGGGGACCTTCGTGGGGCAGAGAC 1500

501 SerCysGlnGlyAspSerGlyGlyProLeuValCysGluGlnAsnAr 517
 |||||
 1501 TCCTGCCAGGAGACAGCGGGGGCTTGTGCTGAGCAGAACACCG 1550
 |||||
 517 gTrpTyrLeuAlaGlyValThrSerTrpGlyThrClyCysGlyGlnArgA 534
 |||||
 1551 CTGGTACCTGGCAGGTGTCACACGCTGGGGCACAGGCTGTGGCCAGAGAA 1600
 |||||
 534 snLysProGlyValTyrThrLysValThrGluValLeuProTrpIleTyr 550
 |||||
 1601 ACAACCTGGTGTGTACACCAAGTGACAGAGTTCCTCCCTGGATTATAC 1650
 |||||
 551 SerLysMetGluSerGluValArgPheArgLysSer 562
 |||||
 1651 AGCAAGATGGAGCGAGGTGCGATTTCATAAAATCC 1686
 |||||

seq_name: /SIDS5/gcdata/geneseq/geneseq-emb1/NA2001A.DAT: AAD05797

seq_documentation_block:

ID AAD05797 standard; DNA; 2067 BP.

XX

AC AAD05797;

XX

DT 31-JUL-2001 (first entry)

XX

DE Human transmembrane serine protease (Endotheliase 2-L) DNA.

XX

KW Human; endotheliase 2-L; protease domain; cytostatic; vulnery; wound;
 KW neutropic; periodontitis; dermatological disorder; gene therapy; scar;
 KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
 KW chronic inflammatory disease; ocular disorder; circulatory disorder;
 KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
 KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;
 KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
 KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;
 KW transmembrane serine protease; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1..2067

FT

FT /tag= a /product= "Human endotheliase 2-L protein"

FT

XX WO200136604-A2.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000WO-0531803.

XX

PR 18-NOV-1999; 99US-0166391.

PR

XX 22-SEP-2000; 2000US-0234840.

XX

PA (CORV-) CORVAS INT INC.

XX

PI Madison EL, Ong EO;

XX

DR WPI: 2001-336001/35.

XX

DR P-PSDB; AAE01944.

XX

PT New nucleic acid encoding a protein comprising endotheliase activity

PT

PT useful in the prevention and treatment of e.g. vascular malformations,

PT

XX cardiovascular disorders, and chronic inflammatory disease -

XX

PS Example 2; Page 139-142; 152pp; English.

XX

CC The present DNA sequence encodes human long form transmembrane serine

CC

CC protease (Endotheliase 2-L) protein.

CC

CC The invention relates to an endotheliase protein, endotheliase protease

CC

CC domain and their corresponding nucleic acid molecules. An endotheliase

CC protein or protease domain of it is useful for the treatment and
 CC diagnosis of disorders associated with aberrant angiogenesis or undesired
 CC neovascularisation. The undesired angiogenesis is associated with
 CC disorders selected from solid neoplasm, vascular malformations and
 CC cardiovascular disorders such as angiofibroma, angiolipoma,
 CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
 CC malformations, haemangiomas and vascular adhesions, dyschondroplasia
 CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
 CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippie Lindau
 CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
 CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
 CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
 CC granulations-burns, hypertrophic scars, liver cirrhosis
 CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
 CC systemic sclerosis and aberrant wound repairs, circulatory disorders
 CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
 CC dyomotility, sclerodactyly and teangiectasis, dermatological disorders
 CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
 CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
 CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
 CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
 CC caused by ocular neovascular disease, corneal graft neovascularisation,
 CC macular degeneration, retinopathy of prematurity, retrolental
 CC fibroplasia and corneal neovascularisation. The nucleic acids of the
 CC invention are also used in gene therapy. The invention also provides
 CC method for screening compounds that modulate angiogenesis.

XX

SQ Sequence 2067 BP; 463 A; 604 C; 540 G; 460 T; 0 other;

alignment_scores:

Quality: 2963.00 Length: 558

Ratio: 5.310 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.283

alignment_block:

US-09-879-792-12 x AAD05797 ..

Align seg 1/1 to: AAD05797 from: 1 to: 2067

1 MetGluArgAspSerHisGlyAsnAlaSerProAlaArgThrProSerAl 17
 |||||
 1 ATGGAGAGGGACAGCCAGGGGAATGCATCTCCAGCAAGAACACCTTCAGC 50
 |||||
 17 aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgA 34
 |||||
 51 TGGAGCATCTCCAGCCAGGCATCTCCAGCTGGGACACTCCAGCGCGGG 100
 |||||
 34 laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
 |||||
 101 CATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGCTGGGACACT 150
 |||||
 51 ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProGly 67
 |||||
 151 CCGGGCGGGCATCTCCAGCCAGGCATCTCCAGCTGGTACACCTCCAGG 200
 |||||
 67 yArgAlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaArgA 84
 |||||
 201 CCGGGCATCTCCAGCGCGGCATCTCCAGCCAGGCATCTCCAGCGCGGG 250
 |||||
 84 laSerProAlaLeuAlaSerLeuSerArgSerSerGlyArgSerSer 100
 |||||
 251 CATCTCCGGCTCTGGCATCACTTCCAGGTCTCTCATCCGGCAGGTCATCA 300
 |||||
 101 SerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeuVa 117
 |||||
 301 TCCGGCCAGTCAAGCTCGGTGACAAACCTCCCAACCAAGAGTGACCTTGT 350
 |||||
 117 lArgAlaThrProValGlyAlaValProIleArgSerSerProAlaArgS 134
 |||||
 351 TAGAGCAACACAGTGGGGCTGTACCCATCCGATCATCTCTCTGCGAGGT 400
 |||||
 134 erAlaProAlaThrArgAlaThrArgGluSerProGlyThrSerLeuPro 150
 |||||
 401 CAGCACCAGCAACAGGCCACAGGGAGAGCCAGGAGTACGAGCTGCC 450
 |||||

151 LysPheThrTrpArgGluGlyGlnLysGlnLeuProLeuIleGlyCysVa 167
|||||
451 AAGTTCACCTGGCGGAGGCGCAGAACAGCTACCGCTCATCGGGTGGT 500
167 LLeuLeuLeuIleAlaLeuValValSerLeuLeuIleLeuPheGlnPhe 184
|||||
501 GCTCCTCCTCATTCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 550
184 rpGlnGlyHisThrGlyIleArgTyrLysGluGlnArgGluSerCysPro 200
|||||
551 GGACGGCCACACAGGATCAGGTACAGGAGGAGGAGGAGGAGGATGCC 600
201 LysHisAlaValArgCysAspGlyValValAspCysLysLeuLysSerAs 217
|||||
601 AAGCACGCTGTTCTGCTGACGGGGTGGTGGACTGCAAGCTGAAGAGTGA 650
217 pGluLeuGlyCysValArgPheAspTyrAspLysSerLeuLeuLysIle 234
|||||
651 CGAGCTGGGCTGCGTGAGGTTTGACTGGGACAGTCTCTGCTTAAATCT 700
234 yrSerGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsn 250
|||||
701 ACTCTGGTCTCTCCATCAGTGGCTTCCCATCTCTAGCAGCACTGGAAT 750
251 AspSerTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHi 267
|||||
751 GACTCCTACTCAGAGAAGACCTGCACGAGCTGGGTTTCGAGAGTGTCTCA 800
267 sArgThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleL 284
|||||
801 CCGACAAACCGAGGTGGCCACAGGATTTTGGCCACAGCTTCTCAATCT 850
284 euArgTyrAsnSerIleGlnGluSerLeuHisArgSerGluCysPro 300
|||||
851 TGAGATACACTCCACCATCCAGGAAGCCCTCCACAGGCTCTGAATGCCCT 900
301 SerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMe 317
|||||
901 TCCAGCGGATATCTCCCTCCAGTGTCCCACTGCGGACTGAGGGCCAT 950
317 tThrGlyArgIleValGlyGlyAlaLeuAlaSerAspSerLysTrpPro 334
|||||
951 GACCGGGCGGATCTGGGAGGGGGCTGGCCCTCGGATAGCAAGTGGCCT 1000
334 rpGlnValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeu 350
|||||
1001 GGCAAGTGAGTCTGCACCTCGGCACCCACATCTGTGGAGGCGCGCTC 1050
351 IleAspAlaGlnTrpValLeuThrAlaAlaHisCysPheValThrAr 367
|||||
1051 ATTGACGCCAGTGGGTGCTCACTGCCGCCCACTGCTTCTTCGTGACCCG 1100
367 gGluLysValLeuGluGlyTyrLysValTyrAlaGlyThrSerAsnLeuH 384
|||||
1101 GGAGAAGGCTCTGGAGGCTGGAGGTTGACGGGACACAGCAACCTGCG 1150
384 isGlnLeuProGluAlaAlaSerIleAlaGluIleIleAlaAsnSerAsn 400
|||||
1151 ACCAGTTCCTTGAGGAGCGCTCCATGTCGAGATCATCATCAACAGCAAT 1200
401 TyrThrAspGluGluAspPyrAspIleAlaLeuMetArgLeuSerLys 417
|||||
1201 TACACCGATGAGGAGGAGCTATGACATCGCCCTCATCGCGCTGTCCAA 1250
417 sProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProMetHisG 434
|||||
1251 GCCCTTGACCTGTCCTGCTCAATCCACCCCTGCTGCTGCCATGCAATG 1300
434 lyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLys 450
|||||
1301 GACAGACCTTTAGCTCAATGAGACCTGCTGGATCACAGGCTTTGGCAAG 1350

451 ThrArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnVa 467
|||||
1351 ACCAGGAGCAGATGACAAAGATCCCTTCTCCGGGAGGTGCAGGT 1400
467 lAsnLeuIleAspPheLysCysAsnAspTyrLeuValTyrAspSert 484
|||||
1401 CAATCTCATCGACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGTT 1450
484 yrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAsp 500
|||||
1451 ACCTTACCCCAAGGATGATGTGTGGGACCTTCGTGGGGGCGAGAGAC 1500
501 serCysGlnGlyAspSerGlyProLeuValCysGluGlnAsnAsnAr 517
|||||
1501 TCCGTCAGGAGACAGCGGGGGCCCTTCTGTCTGTCAGCAGACACACCG 1550
517 gTrpTyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgA 534
|||||
1551 CTGTACTCTGCGAGGTGTACCAAGTGCAGAGGCTGTGGCCAGAGAA 1600
534 snLysProGlyValTyrThrLysValThrGluValLeuProTrpIleTyr 550
|||||
1601 ACAACCTGTGTGTACCAAGTGCAGAAAGTTCCTCCCTGGATTTC 1650
551 SerLysMetGluSerGluValArg 558
|||||
1651 AGCAAGATGGAGAACAGACTCAG 1674
seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF83971
seq_documentation_block:
ID AAF83971 standard; DNA; 1314 BP.
XX
AC AAF83971;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human SER6 nucleotide sequence.
XX
KW Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
KW cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
KW immunosuppressive; antiasthmatic; antifibrinolytic; SER6; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /*tag= a
FT /product= "SER6"
XX
XX WQ200136645-A2.
PN
XX
XX 25-MAY-2001.
PD
XX
XX 17-NOV-2000; 2000WO-US31744.
PF
XX
XX 17-NOV-1999; 99US-0165986.
PR 09-FEB-2000; 2000US-0181347.
PR 03-APR-2000; 2000US-0194195.
PR 05-APR-2000; 2000US-0194839.
PR 07-APR-2000; 2000US-0195637.
PR 13-APR-2000; 2000US-0197080.
PR 15-SEP-2000; 2000US-0232677.
PR 16-NOV-2000; 2000US-0165986.
XX
(CURA-) CURAGEN CORP.
PA
XX
XX Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Taupier R, Rastelli L, Herrmann JU;
XX
XX WPI; 2001-336006/35.
DR
DR P-PSDB; AAB85042.
XX

PT New isolated SERX polynucleotides and polypeptides related to the
PT serine/threonine kinase family of proteins, useful for determining and
PT treating diseases related to altered levels of the polypeptide -
XX
PS
Claim 9; Page 33; 127pp; English.
XX
CC The invention relates to serine/threonine kinase-like proteins, termed
CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
CC via a recombinant expression vector in a host cell in gene therapy
CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
CC polypeptides, polynucleotides, antibodies can also be used to treat or
CC prevent a pathology associated with SERX. Disorders associated with
CC aberrant SERX expression or activity include breast or testicular cancer,
CC cardiovascular defect, autoimmune disorders and asthma. Disorders
CC associated with aberrant SERX-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
CC represents the nucleotide sequence of human SER6.
XX
SQ Sequence 1314 BP; 288 A; 384 C; 376 G; 266 T; 0 other;

alignment_scores:
Quality: 2287.00 Length: 420
Ratio: 5.458 Gaps: 0
Percent Similarity: 99.762 Percent Identity: 99.762
alignment_block:
US-09-879-792-12 x AAF83971 ..
Align seg 1/1 to: AAF83971 from: 1 to: 1314

143 GluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGlnLy 159
4 GAGAGCCAGGTACGAGCCTGCCAAGTTCCACTGGCGGGAGGCCAGAA 53
159 sGlnLeuProLeuGlyCysValLeuLeuLeuLeuLeuLeuValVal 176
54 GCAGCTACCGCTCATCGGGTGGCTGCTCCCTCATTTGGCTGGTGGTT 103
176 erLeuLeuLeuLeuPheGlnPheTrpGlnGlyHisThrGlyIleArgTyr 192
104 CGCTCATCATCTCTTCCAGTTCTGGCAGGGGCCACACAGGATCAGGTAC 153
193 LysGluGlnArgGluSerCysProLysHisAlaValArgCysAspGlyVa 209
154 AAGGACAGAGGAGGAGAGCTCTCCACGACGCTGCTGCTGCGAGGGGT 203
209 lValAspCysLysLeuLysSerAspGluLeuLeuGlyCysValArgPheAspT 226
204 GGTGGACTGCAAGCTGAAGAGTACGAGCTGGGCTGGCTGAGTTTGACT 253
226 rAspLysSerLeuLeuLysIleTyrSerGlySerSerHisGlnTrpLeu 242
254 GGGACAAGTCTCTGTTAAATCTACTCTGGGTCCTCCCATCAGTGGCTT 303
243 ProIleCysSerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGl 259
304 CCCATCTGTAGCAGCAACTGGAATGACTCTCTACTCAGAGAAGACCTGCCA 353
259 nGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAlaHisArgA 276
354 GCAGCTGGGTTCGAGAGTGCTCACCAGACACCGAGGTGGCCACAGGG 403
276 spPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrIleGlnGlu 292
404 ATTTTGGCAACAGCTTCTCAATCTTGAGATACAACTCCACCATCCAGAA 453
293 SerLeuHisArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCy 309
454 AGCCTCCACAGGCTGAATGCCCTTCCACGCGGTATATCTCCCTCCAGTG 503
309 sSerHisCysGlyLeuArgAlaMetThrGlyArgIleValGlyGlyAlaL 326

|||||
504 TTCCACTGCGGACTGAGGCGCATGACCGGCGGATCGTGGAGGGCGC 553
326 euAlaSerAspSerLysTrpTrpGlnValSerLeuHisPheGlyThr 342
|||||
554 TGGCTCGGATAGCAAGTGGCTTGGCAAGTGAGTCTGCACCTCGGCACC 603
343 ThrHisLeuCysGlyGlyThrLeuLeuAspAlaGlnTrpValLeuThrAl 359
|||||
604 ACCCACATCTCTGGAGGCACGCTCATTTGACGCCAGTGGGTGCTCAGTCG 653
359 aAlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLysV 376
|||||
654 GCGCCACTGCTTCTGCTGAGCCCGGAGAGGTCTTGAGGGCTGGGAAGG 703
376 alTyrAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIle 392
|||||
704 TGTACGGCGGCACCGCAACCTGCACCATGTCCTGAGGCAGCCTCCATT 753
393 AlaGluLeuIleLeuAsnSerAsnTyrThrAspGluGluAspAspTyrAs 409
754 GCCGAGATCATCATCAACAGCAATTACACCGATGAGGAGGAGCATATGA 803
409 pIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleH 426
|||||
804 CATGCCCTCATGCGGCTGTCCAAAGCCCCCTGACCTGTCCGCTCACATCC 853
426 isProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsnGluThr 442
854 ACCCTGCTGCTCCCATGATGATGAGCAGACCTTTAGCCTCAATGAGACC 903
443 CysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLysThrSe 459
904 TGCTGGATCACAGGCTTTGGCAAGCAGGAGGAGAGATGACAAAGACATC 953
459 rProPheLeuArgGluValGlnValAsnLeuLeuIleAspPheLysCysA 476
954 CCCCTCTCTCGGGAGGTGTCAGGTCAATCTCATCGACTTCAAGAAATGCA 1003
476 snAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAla 492
1004 ATGACTACTGGTCTATGACAGTTACCTTACCCCAAGGATGATGTGTCT 1053
493 GlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyPr 509
1054 GGGACCTTCTGTTGGGGCAGAGACTCTCTGCCAGGAGAGACAGCGGGGCC 1103
509 oLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSerT 526
1104 TCTTGTCTGTAGCAGAAACACCGCTGGTACCTGGCAGGTGTCACAGCT 1153
526 rpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysVal 542
1154 GGGCAGACGCTGTGGCCAGAGAAACAACTGGTGTGTACACCAAGTG 1203
543 ThrGluValLeuProTrpIleTyrSerLysMetGluSerGluValArgPh 559
1204 ACAGAAGTCTCTCCCTGGATTTACAGCAAGATGGAGAGCGAGGTGCGATT 1253
559 eArgLysSer 562
1254 CACAAAATCC 1263

seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF83973
seq_documentation_block:
ID AAF83973 standard; DNA; 1314 BP.
XX
AC AAF83973;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of SER8, a partial reverse complement of SER6.

XX Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
 KW cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
 KW immunosuppressive; antiasthmatic; antifibrinolytic; SER6; SER8; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200136645-A2.
 XX
 PD 25-MAY-2001.
 XX
 PF 17-NOV-2000; 2000WO-US31744.
 XX
 PR 17-NOV-1999; 9905-0165986.
 PR 09-FEB-2000; 2000US-0181347.
 PR 03-APR-2000; 2000US-0194195.
 PR 05-APR-2000; 2000US-0194839.
 PR 07-APR-2000; 2000US-0195637.
 PR 13-APR-2000; 2000US-0197080.
 PR 15-SEP-2000; 2000US-0232677.
 PR 16-NOV-2000; 2000US-0165986.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
 PI Taupier R, Rastelli L, Herrmann JL;
 XX
 DR WPI: 2001-336006/35.
 XX
 PT New isolated SERX polynucleotides and polypeptides related to the
 PT serine/threonine kinase family of proteins, useful for determining and
 PT treating diseases related to altered levels of the polypeptide -
 XX
 PS Claim 9; Page 38; 127pp; English.
 XX
 CC The invention relates to serine/threonine kinase-like proteins, termed
 CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
 CC via a recombinant expression vector in a host cell in gene therapy
 CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
 CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
 CC polypeptides, polynucleotides, antibodies can also be used to treat or
 CC prevent a pathology associated with SERX. Disorders associated with
 CC aberrant SERX expression or activity include breast or testicular cancer,
 CC cardiovascular defect, autoimmune disorders and asthma. Disorders
 CC associated with aberrant SERX-8 expression include, blood clotting
 CC hematopoietic and tumor related disorders. The present sequence
 CC represents the nucleotide sequence of human SER8, a partial reverse
 CC complement of SER6.
 XX
 SQ Sequence 1314 BP; 266 A; 376 C; 384 G; 288 T; 0 other;

alignment_scores:
 Quality: 2287.00 Length: 420
 Ratio: 5.458 Gaps: 0
 Percent Similarity: 99.762 Percent Identity: 99.762
 alignment_block:
 US-09-879-792-12 x AAF83973/rev ..
 Align seg 1/1 to reverse of: AAF83973 from: 1 to: 1314
 143 GluSerProGlyThrSerLeuProLysPheThrTrpArgGluGlyGlnLy 159
 |||||
 1311 GAGAGCCAGGTACGAGCCCTGCCAAGTTCCACCTGGCGGAGGCCAGAA 1262
 159 sGlnLeuProLeuIleGlyCysValLeuLeuLeuAlaLeuValVal 176
 |||||
 1261 GCAGCTACCCCTCATCGGGTGGCTGCTCCCTCATTTGCCCTGGTGGTTT 1212
 176 erLeuIleLeuLeuPheGlnPheTrpGlnGlyHisThrGlyIleArgTyr 192
 |||||
 1211 CGCTCATCTCTCTTCCAGTTCTGGCAGGGCCACACAGGGATCAGGTAC 1162

193 LysGluGlnArgGluSerCysProLysHisAlaValArgCysaspGlyVa 209
 |||||
 1161 AAGGAGCAGAGGAGAGCTGTCCCAAGCAGCTGTTCGCTGTGACGGGT 1112
 209 lValAspCysLysLeuLysSerAspGluLeuGlyCysValArgPheAspt 226
 |||||
 1111 GGTGGACTGCAAGCTGAAGAGTGACGAGCTGGGCTGCTGAGGTTTGACT 1062
 226 rAspLysSerLeuLysIleTyrSerGlySerSerHisGlnTrpLeu 242
 |||||
 1061 GGGACAAGTCTCTGCTTAAATCTACTCTGGGTCTCTCCCATCAGTGGCTT 1012
 243 ProIleCysSerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGl 259
 |||||
 1011 CCCATCTGTAGCAGCAACTGGAATGACTCTCTACTCTCAGAGAAGACCTGCCA 962
 259 nGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAlaHisArgA 276
 |||||
 961 GCAGCTGGGTTTCGAGAGTGTCTACCGGACAACCGAGGTGCCCCACAGG 912
 276 spPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrIleGlnGlu 292
 |||||
 911 ATTTGGCCAAACAGCTTCTCAATCTTGAGATACAACCTCCACATCCAGGAA 862
 293 SerLeuHisArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCy 309
 |||||
 861 AGCTCCACAGGTCTGAATGCCCTTCCACGGGTATATCTCCCTCCAGTG 812
 309 sSerHisCysGlyLeuArgAlaMetThrGlyArgIleValGlyGlyAlaL 326
 |||||
 811 TTCCCATCGGGACTGAGGGCCATGACCGGGCGGATCTGGAGGGGCGC 762
 326 euAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThr 342
 |||||
 761 TGGCTCGGATAGCAAGTGGCTTGGCAAGTAGCTGCACTTCGGGCACC 712
 343 ThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAl 359
 |||||
 711 ACCACATCTGTGGAGCAGCGCTCATTTGAGCCCGCAGTGGGTCTCACTGC 662
 359 aAlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLysV 376
 |||||
 661 CGCCCATCTCTCTTCGTGACCCGGGAGAGGTCTCTGGAGGGCTGGAAGG 612
 376 alTyrAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIle 392
 |||||
 611 TGTACGGCGGCACAGCAACCTGCACCAGTTGCCTGAGGCAGCCTCCATT 562
 393 AlaGluIleIleIleAsnSerAsnTyrThrAspGluAspAspTyrAs 409
 |||||
 561 GCGGAGATCATCATCAACAGCAATTACACCGATGAGGAGGACGACTATGA 512
 409 pIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleH 426
 |||||
 511 CATCGCCCTCATCGGGCTGTCCAGAGCCCTTGACCCCTGTCCGCTCACATCC 462
 426 isProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsnGluThr 442
 |||||
 461 ACCCTGCTTGCCTCCCATCATGATGGACAGACCTTTAGCCTCAATGAGACC 412
 443 CysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLysThrSe 459
 |||||
 411 TGCTGGATCACAGCTTTGGCAAGACCAGGAGACAGATGACAAGACATC 362
 459 rProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysCysA 476
 |||||
 361 CCCCTTCTCCGGAGGTGAGGCTCAATCTCATCGACTTCAAGAAATGCA 312
 476 snAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAla 492
 |||||
 311 ATGACTACTTGGTCTATGACAGTTACCTTACCCCAAGGATGATGTGTGCT 262

493	GlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyPr	509
261	GGGGACCTTCGTGGGGCAGAGACTCCTGCCAGGGAGACAGCGGGGGCC	212
509	oLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSerT	526
211	TCTGTGCTGTGAGCAGAACCGCTGGTACCTGGCAGGTGTCCACGACT	162
526	rpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysVal	542
161	GGGGCACAGGCTGTGGCCAGAGAAACAAACCTGGTCTGTACACCAAGTG	112
543	ThrGluValLeuProTrpIleTyrSerLysMetGluSerGluValArgPh	559
111	ACAGAAGTTCCTCCCTGGATTTACAGCAAGATGGAGAGCGAGTGGGATT	62
559	eArgLysSer	562
61	CACAAATCC	52
seq_name: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH14850		
seq_documentation_block:		
XX	AAH14850 standard; cDNA; 2192 BP.	
XX	AAH14850;	
XX		
DT	26-JUN-2001 (first entry)	
XX		
DE	Human cDNA sequence SEQ ID NO:12681.	
XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene ti	
XX		
OS	Homo sapiens.	
XX		
PN	EP1074617-A2.	
XX		
PD	07-FEB-2001.	
XX		
PF	28-JUL-2000; 2000EP-0116126.	
XX		
PR	29-JUL-1999; 99JP-0248036.	
PR	27-AUG-1999; 99JP-0300253.	
PR	11-JAN-2000; 2000JP-0118776.	
PR	02-MAY-2000; 2000JP-0183767.	
PR	09-JUN-2000; 2000JP-0241899.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX		
DR	WPI; 2001-318749/34.	

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB302416 to
CC AAB98993 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
SQ Sequence 2192 BP; 531 A; 655 C; 536 G; 470 T; 0 other;

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in

```
|||||
691 AGGAGAGCTGTCCCAAGCAGCGTGTTCGCTGTGACGGGTGGTGACTG 740
212 sLysLeuLysSerAspGluLeuGlyCysValArgPheAspTrpAspLys 229
|||||
741 CAAGCTCAAGAGTGACGAGCTGGCTGGCTGAGGTTTGACTGGGACAAGT 790
229 erLeuLeuLysIleTyrSerGlySerSerHisGlnTrpLeuProIleCys 245
|||||
791 CTCTGCTTAAATCTACTCTGGGTCTCCCATCATCAGTGGCTTCCCATCTGT 840
246 SerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGlnLeuG1 262
|||||
841 AGCAACACTGGGAATGACTCTCTACTAGAGAAGACCTGCCAGCAGCTGGG 890
262 yPheGluSerAlaHisArgThrThrGluValAlaHisArgAspPheAlaA 279
|||||
891 TTTTCGAGAGTGCTCACCGGACACCGAGGTGCCACAGGATTTTCCCA 940
279 snSerPheSerIleLeuArgTyrAsnSerThrIleGlnGluSerLeuHis 295
|||||
941 ACAGCTTCTCAATCTTGAGATACAACCTCCACCATCCAGGAAAGCTCCAC 990
296 ArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCysSerHisCy 312
|||||
991 AGGTCTGAATGGCCCTTCCCAAGCGTATATCTCCCTCCAGTGTCCCACTG 1040
312 sGlyLeuArgAlaMetThrGlyArgIleValGlyGlyAlaLeuAlaSerA 329
|||||
1041 CGGACTGAGGCGCATGACCGGCGGATCTGTGGAGGGCGCTGCCCTCGG 1090
329 spSerTyrTrpProTrpGlnValSerLeuHisPheGlyThrThrHisIle 345
|||||
1091 ATAGCAAGTGGCCCTTGCAAGTAGTGTGACCTTCGCGCACCCACCATC 1140
346 CysGlyGlyThrIleuLeuAspAlaGlnTrpValLeuThrAlaAlaHisCy 362
|||||
1141 TGTGGAGGCACGCTCAFTAGCGCCAGTGGGTGCTCACTGCCGCCCACTG 1190
362 sPhePheValThrArgGluLysValLeuGluGlyTrpLysValTyrAlaG 379
|||||
1191 CTTCCTTCGTACCCGGGAGAGGCTCTGGAGGCTGGAAGGTGTACGCGG 1240
379 lyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIleAlaGluIle 395
|||||
1241 GCACGACAACTGCACACAGTGTGCTGAGGAGCCTCCATTGCGGAGATC 1290
396 IleIleAsnSerAsnTyrThrAspGluGluAspTyrAspIleAlaLe 412
|||||
1291 ATCATCAACAGCAATTACACCGATGAGGAGGAGGACATATGACATCGCCCT 1340
412 uMetArgLeuSerLysProLeuThrLeuSerAlaHis..... 424
|||||
1341 CATCGCGCTCTCCAAGCCCTGACCCCTGTCCGGTGAGGGAATCTGCACCTC 1390
425 ..IleHisProAlaCysLeuProMethis 433
|||||
1391 CCCGCTCTCTGCCGCCCGCCAGCAGCAC 1419
```

seq_name: /SID5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF83972
seq_documentation_block:
ID AAF83972 standard; DNA; 1078 BP.

XX
AC
XX
DT
XX
DE
KW
KW
KW

AAF83972;
06-AUG-2001 (first entry)
Nucleotide sequence of SER7, a partial reverse complement of SER6.
Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
immunosuppressive; antiasthmatic; antifibrinolytic; SER6; SER7; ss.

XX Homo sapiens.
XX WO200136645-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31744.
XX
PR 17-NOV-1999; 99US-0165986.
PR 09-FEB-2000; 2000US-0181347.
PR 03-APR-2000; 2000US-0194195.
PR 05-APR-2000; 2000US-0194839.
PR 07-APR-2000; 2000US-0195637.
PR 13-APR-2000; 2000US-0197080.
PR 15-SEP-2000; 2000US-0232677.
PR 16-NOV-2000; 2000US-0165986.
XX (CURA-) CURAGEN CORP.
PA
XX Quinn KB, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Taupier R, Rastelli L, Herrmann JL;
XX
DR WPI; 2001-336006/35.
XX
PT New isolated SERX polynucleotides and polypeptides related to the
PT serine/threonine kinase family of proteins, useful for determining and
PT treating diseases related to altered levels of the polypeptide -
XX
PS Claim 9; Page 37; 127pp; English.
XX
CC The invention relates to serine/threonine kinase-like proteins, termed
CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
CC via a recombinant expression vector in a host cell in gene therapy
CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
CC polypeptides, polynucleotides, antibodies can also be used to treat or
CC prevent a pathology associated with SERX. Disorders associated with
CC aberrant SERX expression or activity include breast or testicular cancer,
CC cardiovascular defect, autoimmune disorders and asthma. Disorders
CC associated with aberrant SER5-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
CC represents the nucleotide sequence of human SER7, a partial reverse
CC complement of SER6.
XX
SQ Sequence 1078 BP; 221 A; 296 C; 319 G; 242 T; 0 other;

alignment_scores:
Quality: 1869.00 Length: 342
Ratio: 5.481 Gaps: 0
Percent Similarity: 99.708 Percent Identity: 99.708
alignment_block:
US-09-879-792-12 x AAF83972/rev ..
Align seg 1/1 to reverse of: AAF83972 from: 1 to: 1078
221 CysValArgPheAspTrpAspLysSerLeuLeuLysIleTyrSerGlyse 237
|||||
1077 TCGGTGAGGTTTGACTGGACAAAGTCTCTGCTTAAATCTACTCTGGGTC 1028
237 rSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrS 254
|||||
1027 CTCCATCAGTGGCTTCCCATCTCTAGCAGCACTGGAATGACTCTACT 978
254 erGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThr 270
|||||
977 CAGAGAAGACTGCCAGCAGCTGGGTTTCGAGAGTGCTCCACGGACAACC 928
271 GluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrAs 287
|||||
927 GAGGTTCCCCACAGGGATTTTGCCAACACAGCTTCTCAATCTTGAGATACAA 878

```
287 nSerThrIleGlnGluSerLeuHisArgSerGluCysProSerGlnArgT 304
|||||
877 CTCACCATCCAGAAAGCCCTCCACAGGCTGTAATGCCTTCCAGCGGT 828
|||||
304 YrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMetThrGlyArg 320
|||||
827 ATATCTCCCTCCAGGTGTCCACACTGCCGACTGAGGGCCATGACCGGCGG 778
|||||
321 IleValGlyGlyAlaLeuAlaSerAspSerLysTrpProTrpGlnValse 337
|||||
777 ATCGTGGAGGGCGCTGGCTCGGATAGCAAGTGGCTTGGCAAGTGAG 728
|||||
337 rLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuIleAspIaG 354
|||||
727 TCTGCACCTTCGGACCAACCCACATCTCTGGAGCAGCAGCTCATTTGACGCC 678
|||||
354 lnrTrpValLeuThrAlaAlaHisCysPhePheValThrArgGluLysVal 370
|||||
677 AGTGGGTGCTCACTGCCGCCACCTGCTTCTTGTGACCCGGGAGAGGTC 628
|||||
371 LeuGluGlyTrpLysValTyAlaGlyThrSerAsnLeuHisGlnLeuPr 387
|||||
627 CTGGAGGGCTGGAAGGTGTACGGGGCACCAGCAACCTGCACCAGTTGCC 578
|||||
387 oGluAlaAlaSerIleAlaGluIleIleIleAsnSerAsnTyrThrAspG 404
|||||
577 TGAGGCAGCCTCCATGTCGCCAGATCATCATCAACAGCAATTACACCGATG 528
|||||
404 luGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysProLeuThr 420
|||||
527 AGAGGAGCACTATGACATCGCCCTCATGCGGCTGTCCAAAGCCCTGACC 478
|||||
421 LeuSerAlaHisIleHisProAlaCysLeuProMethHisGlyGlnThrPh 437
|||||
477 CTGTCCGCTCACATCCACCTGCTTGCTCCCTCCCATGATGCAGACACCTT 428
|||||
437 eSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluT 454
|||||
427 TAGCCCTCAATGAGACCTGCTGGATCATCAGGCTTTGGCAAGACAGGAGA 378
|||||
454 hrAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuIle 470
|||||
377 CACATGACAGACATCCCTCTCTCGGGAGGTGCAGGTCAATCTCATC 328
|||||
471 AspPheLysLysCysAsnAspTyrLeuValTyAspSerTyrLeuThrPr 487
|||||
327 GACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGATTACCTTACCCC 278
|||||
487 oArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnG 504
|||||
277 AAGGATGATGTGTCTGGGACCTTCTGTGGGGCAGAGACTCTCGCAGG 228
|||||
504 lyAspSerGlyGlyProLeuValCysGluGlnAsnAsnArgTrpTyrLeu 520
|||||
227 GAGACAGCGGGGGCGCTCTGTCTGTGAGCAGAACACCCCTGGTACCTG 178
|||||
521 AlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProG1 537
|||||
177 GCAGGTGTCAACAGCTGGGGCAGAGGCTGTGGCCAGAGAAACAACCTGG 128
|||||
537 yValTyThrLysValThrGluValLeuProTrpIleTyrSerLysMetG 554
|||||
127 TGTGTACACCAAGTGACAGAAGTCTTCCCTGGATTTCACAGCAAGATGG 78
|||||
554 luSerGluValArgPheArgLysSer 562
|||||
77 AGAGCGAGGTGCCATTCACAAAATCC 52
```

seq_name: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: AAD02322
seq_documentation_block:
ID AAD02322 standard; cDNA; 1222 BP.

```
XX AAD02322;
XX 28-MAR-2001 (first entry)
XX Human serine protease #3, encoding HATEE38 cDNA clone.
XX
KW Human; serine protease; osteopathic; immunosuppressive; antiallergic;
KW antiinflammatory; cytostatic; cardiant; neuroprotective; nootropic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 76..1194
/*tag= a
/*product= "Human serine protease #3 from clone HATEE38"
FT 76..963
/*tag= b
FT /*product= "Human serine protease #3 fragment from
FT clone HATEE38"
FT /*note= "Does not include stop codon"
FT /*partial
XX
PN WO200068247-A2.
XX
XX 16-NOV-2000.
XX
XX 05-MAY-2000; 2000WO-US12207.
XX
XX 07-MAY-1999; 99US-0133239.
PR 20-MAY-1999; 99US-0135163.
PR 03-AUG-1999; 99US-0147005.
PR 09-SEP-1999; 99US-0152935.
PR 01-NOV-1999; 99US-0162979.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Shi Y, Young PE, Ni J;
XX
XX WPI: 2000-679799/66.
XX P-PSDB; AAY72092, AAY72108.
XX
XX New nucleic acid molecules encoding human serine protease polypeptides,
XX useful for diagnosis, prevention and/or treatment of disorders e.g.
XX osteoporosis, lupus erythematosus and Alzheimer's -
XX
XX Claim 1; Page 271-272; 289pp; English.
XX
XX The present cDNA sequence encodes human serine protease #3 from clone
XX HATEE38 (ATCC Deposit No: PTA27).
XX The invention relates to human serine proteases and their cDNA clones.
XX It is used in methods for the diagnosis, prevention and treatment of
XX various disorders related to serine protease such as bone formation
XX disorders (osteoporosis), connective tissue disorders (arthritis),
XX autoimmune disorders (systemic lupus erythematosus), wound healing, male
XX reproductive system disorders (testicular cancer), digestion and food
XX absorption disorders (Crohn's disease), neurodegenerative diseases
XX (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
XX proliferative and cancerous conditions (acute myelogenous leukaemia),
XX allergic reactions (asthma), cardiovascular disorders (arrhythmia),
XX ocular disorders (glaucoma) and infectious diseases caused by bacteria,
XX viruses, fungi or parasites. It is also useful for screening therapeutic
XX compounds. Serine proteases are used as immunological probes or
XX polymorphic markers for the identification of chromosomes, cells and
```

CC tissues in biological samples, identification of male contraceptive
CC agents, delivery of compositions to targetted cells expressing a
CC receptor for serine protease, hybridisation probes and molecular weight
CC markers. Serine protease nucleic acids are also useful in gene therapy.
XX
SQ Sequence 1222 BP; 303 A; 314 C; 305 G; 300 T; 0 other;

alignment_scores:

Quality: 1436.00 Length: 264
Ratio: 5.439 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.485

alignment_block:

US-09-879-792-12 x AAD02322 ..

Align seg 1/1 to: AAD02322 from: 1 to: 1222

295 HisArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCysSerHi 311
|||||
10 CACAGGTCTGAATGCCCTTCCACGGGTATATCTCCCTCCAGTGTCCCA 59
311 sCysGlyLeuArgAlaMetThrGlyArgIleValGlyAlaLeuAla 328
|||||
60 CTGCGGACTGAGGGCCCATACCGGGGGATCGTGGGAGGGCGCTGGCCT 109
328 erAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThrHis 344
|||||
110 CGGATAGCAAGTGGCCTTGGCAGTGAAGTGCACATCGGCACCCAC 159
345 IleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaLaHi 361
|||||
160 ATCTGTGGAGGCACGGCTCATTCATTCAGCCCGGCTGCTCCTCCCGCCA 209
361 sCysPhePheValThrArgGluLysValLeuGluGlyTrpLysValTyrA 378
|||||
210 CTGCTCTCTGTCACCGGAGAGGAGGCTGGAGGGCTGAAGGTGTACG 259
378 laGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerIleAlaGlu 394
|||||
260 CGGCGACCAACCTGCACCACTGCCTGAGGAGGAGGCTCCATTCGGAG 309
395 IleIleIleAsnSerAsnTyrThrAspGluGluAspAspTyrAspIleAl 411
|||||
310 ATCATCATCAACAGCAATTACACCGATGAGGAGGAGCATATGACATCGC 359
411 aLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleHisProA 428
|||||
360 CCTCATGGCGCTGTCCNAGCCCTGACCCCTGTCGGTCCATCCACCCCTG 409
428 laCysLeuProMetHisGlyGlnThrPheSerLeuAsnGluThrCysTrp 444
|||||
410 CTTGCGCTCCCATGATGGACAGACCTTTAGCCTCAATGAGACCTGCTGG 459
445 IleThrGlyPheGlyLysThrArgGluThrAspAspLysThrSerProPh 461
|||||
460 ATCACAGGCTTTGGCAAGACCGAGGAGACAGATGACAAGACATCCCCCTT 509
461 eLeuArgGluValGlnValAsnLeuIleAspPheLysLysCysAsnAspT 478
|||||
510 CTTCCGGGAGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACT 559
478 yrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAlaGlyAsp 494
|||||
560 ACTTGGTCTATGACAGTTACCTTACCCCAAGGATGATGTGTGCTGGGAC 609
495 LeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyProLeuVal 511
|||||
610 CTTGCTGGGGGAGAGACTCCTGCCAGGAGAGACAGCGGGGGGCTCTTGT 659
511 lCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSerTrpGlyT 528
|||||
660 CTGTGAGCAGACAACCGCTGTGACCTGGCAGGTGTCCACAGCTGGGCA 709

528 hrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysValThrGlu 544
|||||
710 CAGGCTGTGGCCAGAGAAACAACCTGGTGTGTACACCAAAAGTGACAGAA 759
545 ValLeuProThrPileTyrSerLysMetGluSerClnuValArg 558
|||||
760 GTTCTTCCCTGGATTACAGCAAGATGGAGAACAGAGCTCAG 801

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH07577

seq_documentation_block:

ID AAH07577 standard; cDNA; 852 BP.

XX

AC AAH07577;

DT 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:4412.

DE Human; primer; detection; diagnosis; antisense therapy; ss.

KW Homo sapiens.

XX EP1074617-A2.

PN 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs .

PS Claim 1; SEQ ID 4412; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAH32446 to
AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.

XX

Seq Sequence 852 BP; 180 A; 284 C; 227 G; 153 T; 8 other;

alignment_scores:

Quality: 960.00 Length: 246
Ratio: 4.344 Gaps: 13
Percent Similarity: 89.837 Percent Identity: 89.024

alignment_block:

US-09-879-792-12 x AAH07577 ..

Align seg 1/1 to: AAH07577 from: 1 to: 852

```
1 MetGluArgAspSerHisGlyAsnAlaSerProAlaArgThrProSerAl 17
|||||
91 ATGAGAGGACAGCCACCGGATGTCATCTCCAGCAAGAACACCTTCAGC 140
|||||
17 aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgA 34
|||||
141 TGGAGCATCTCCAGCCAGGCATCTCCAGCTGGGACACCTCCAGGCGGG 190
|||||
34 laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
|||||
191 CATCTCAGCCAGGATCTCCAGCCAGGCATCTCCAGCTGGGACACCT 240
|||||
51 ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProGly 67
|||||
241 CCGGGCCGGCATCTCCAGCCAGGCATCTCCAGCTGGTACACCTCCAGG 290
|||||
67 yArgAlaSerProGlyArg.....AlaSerProAlaGlnA 79
|||||
291 CCGGGCATCTCCAGGCGGCATCTCCAGCCAGGCATCTCCAGCCAGG 340
|||||
79 laSerProAlaArgAlaSerProAlaLeuAlaSerLeuSerArgSer 95
|||||
341 CATCTCAGCCGGGATCTCCGGCTCTGGCATCTTCCAGGCTCTCA 390
|||||
96 SerGlyArgSerSerAlaArgSerAlaSerValThrThrSerProTh 112
|||||
391 TCCGGCAGGTCATCTCCGCGAGTCAGCTCGTGACAACTCCCAAC 440
|||||
112 rArgValThrLeuValArgAlaThrProValGlyAlaValProIleArgS 129
|||||
441 CAGAGTGATCTGTAGAGCAACACCACTGGGGGCTGTACCATCCGAT 490
|||||
129 erSerProAlaArgSerAlaPro. AlaThrArgAlaThrArgGluSerPr 145
|||||
491 CATCTCTCCAGTCAGCAAGTTCACCTGGCGGGAGGCCCAAGAGCAGCT 540
|||||
145 oGlyThr. SerLeuProLysPheThrTrpArgGluGlyGln. LysGlnLe 161
|||||
541 AGGTACGAAGCCTGCCAAGTTCACCTGGCGGGAGGCCCAAGAGCAGCT 590
|||||
161 uProLeuIleGlyCysValLeuLeuLeuIleAlaLeuValValSerLeuI 178
|||||
591 ACCGCTCATCGGTGGTGTCTCTCTCATFTGCCCTGGTGTTCGCTCA 640
|||||
178 leIleLeuPheGlnPheThrGlnGlyHisThrGlyIleArgTyrLys.G1 194
|||||
641 TCATCTCTTCCAGTCTTGGCANGGCCACACAGGGATNAGGTACAGGGA 690
|||||
194 uGlnArg. GluSerCys. ProLysHisAlaValArgCys.... AspGlyVa 209
|||||
691 GCAAGGGGAGAGCTGTCCAAAGCAGCGTGTTCGCTGNTGAACGGGT 740
|||||
209 l.valAspCysLysLeu. LysSer. AspGluLeu.... GlyCysValArg. 223
|||||
741 GGGTGACGTCAAGCTGGAGAGCTTGACNNACCTGGGGGCTGCTTTANG 790
|||||
224 ..PheAspTrpAspLysSer 229
|||||
791 GTTTTCACTGGGAAAAAGT 810
```

seq_name: /SIDSS/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AAS91044

seq_documentation_block:

ID: AAS91044 standard; cDNA; 1212 BP.

XX AAS91044;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #26848.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG26857.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1; SEQ ID No 26848; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1212 BP; 275 A; 348 C; 328 G; 261 T; 0 other;

alignment_scores:

Quality: 819.00 Length: 409
Ratio: 3.545 Gaps: 15
Percent Similarity: 56.479 Percent Identity: 49.144

alignment_block:

US-09-879-792-12 x AAS91044 ..

Align seg 1/1 to: AAS91044 from: 1 to: 1212


```

|||||
1465 AGTGGAGGCGCTGGTCAATCGAAGAACAAATATCTGGTGGCTGATAGG 1514
522 yValThrSerTyrGlyCysGlyGlnArgAsnLysProGlyValT 539
|||||
1515 GGATACAAAGCTGGGTCTGGCTGTGCAAGCTTACAGACCAGAGTGT 1564
539 yThrLysValThrGluValLeuProTyrLysMetGluSer 555
|||||
1565 ACGGAATGTGATGTTATTCACGACTGGATTATTCGACAAATGAGGCA 1614
556 Glu 556
1615 GAC 1617

```

seq_name: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAS64178

seq_documentation_block:

ID AAS64178 standard; cDNA; 3245 BP.

AC AAS64178;

DT 29-JAN-2002 (first entry)

DE Human prostate cDNA sequence #597.

KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.

OS Homo sapiens.

PN WO200173032-A2.

PD 04-OCT-2001.

PF 27-MAR-2001; 2001WO-US09919.

PR 27-MAR-2000; 2000US-0536857.

PR 09-MAY-2000; 2000US-0568100.

PR 12-MAY-2000; 2000US-0570737.

PR 13-JUN-2000; 2000US-0593793.

PR 27-JUN-2000; 2000US-0605783.

PR 10-AUG-2000; 2000US-0636215.

PR 29-AUG-2000; 2000US-0651236.

PR 06-SEP-2000; 2000US-0657279.

PR 02-OCT-2000; 2000US-0679426.

PR 10-OCT-2000; 2000US-0685166.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

PI Panger GR, Retter MW, Stolk JA, Day CH, Vedwick TS, Carter D;

PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;

XX WPI; 2001-639232/73.

DR P-PSDB; AAU69960.

XX New human prostate-specific polypeptides and polynucleotides useful for

PT the diagnosis and treatment of cancer, especially prostate cancer -

PS Claim 1; Page 570-571; 579pp; English.

XX The invention relates to isolated prostate-specific

CC polynucleotides, polypeptides, fusion proteins of the polypeptides,

CC antibodies raised against the polypeptides (or antigenic epitopes

CC derived from them) and antigen-presenting cells expressing the

CC polypeptides. The antibodies are useful for detecting the presence of

CC cancer, especially prostate cancer. The polypeptides, polynucleotides and

CC the antigen-presenting cells are useful for stimulating and/or expanding

CC T cells specific for a tumour protein, and for inhibiting the development

CC of cancer especially prostate cancer. Compositions comprising the

CC polynucleotide and/or polypeptide are useful for stimulating an immune

CC response, and for treating cancer. The oligonucleotide is useful for

CC detecting cancer. The present sequence is a prostate specific

```

CC polynucleotide of the invention.
XX
SQ Sequence 3245 BP; 774 A; 818 C; 870 G; 783 T; 0 other;

alignment_scores:
  Quality: 743.00      Length: 551
  Ratio: 2.173        Gaps: 17
  Percent Similarity: 62.069  Percent Identity: 33.394

alignment_block:
US-09-879-792-12 x AAS64178 ..
Align seg 1/1 to: AAS64178 from: 1 to: 3245

43 GlnAlaSerProAlaGlyThrProGlyArgAlaSerProAlaGlnAl 59
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 CGCGGAGCTAAGCAGGCGGAGGCGGAGGCGGAGGCGGCGGCGG 65
59 aSerProAlaGlyThrProGlyArgAlaSerProGlyArg..... 73
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 GAGCGCGCCTGGAGCGCGCGCATATTGAAACATTCCAGATACCTAT 115
74 .....AlaSerProAlaGlnAlaSerProAlaArgAla 84
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 CATTACTCGATGCTGTGATAACAGCAAGATGGCTTGAACCTCAGGTCA 165
85 SerProAlaAlaSerLeuSerArgSerSerSerGlyArgSerSer 101
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166 CCACCAGCTATTGGACCTTACTATGAAACCATGGATACCAACCGGAAA 215
101 rAlaArgSerAlaSerValThrThrSerProThrArgValThrLeuVal 118
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 CCCTATCCGCGCACAGCCCACTGTGTCCCACT...GTCTACGAGGTGC 262
118 rg.....AlaThrProValGlyAlaValProIleArgSer 129
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 ATCCGGCTCAGTACTACCCGTCGCCCGTCCCGCCAGTAGTACGCCCGAGGTC 312
130 SerProAlaArgSerAlaProAlaThrArgGlySerPro... 145
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 CTGAGCGAGGCTTCCAAACCCGCTGCTGCAGCAGCCCAATCCCATC 362
146 GlyThrSerLeuProLysPheThrTrpArgGlyGlnLysGlnLeuP 162
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
363 CGGGACAGTGTGCACCTCAAGACT.....AAGAAAGCACTGT 400
162 roLeu.....IleGlyCysValLeuLeuLeuIleAlaLeuValVal 175
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 GCATCACCCTTGACCTGGGACCTTCTCTGGGAGCTGGCTGGCGGCT 450
176 SerLeuIleLeuPheGlnPheTrpGln.....G1 186
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 GGCCTACTC.....TGGAGTTTCATGGCGCAGCAAGTGCTC 485
186 yHisThrGlyLeuArgTyrLysGluGlnArgGlySerCysProLysHis 203
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
486 CAATCTGGGATAGAGTGC...GACTCTCAGGTACCTGTCATCAACCCCT 532
203 laValArgCysAspGlyValValAspCysLysLeuLysSerAspGluLeu 219
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
533 CTAACTGGTGTGATGGCTGTACACTGCCCGCGGGAGGAGGAGAAAT 582
220 GlyCysValArgPheAspTrpAspLysSerLeuLysIleTyrSerG1 236
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
583 CGGTGTCTTCCTCTACGGATCAAACTTCTCAGGTGTACTCATC 632
236 ySerSerHisGlnTrpLeuProLysSerSerAsnTrpAsnAspSer 253
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
633 TCAGAGGAAGTCTCTGGCACCTGTGTGCCAAGACGACTGGAACGAGAA 682
253 yrSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThr 269
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```
683 ACGGGCGGGCGGCTGCAGGACATGGGCTATAGAAATAATTTTACTCT 732
270 ThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTy 286
      : : : : : ||| : : : : : ||| : : : : :
733 AGCAA...GGAATAGTGATGACAGGGATCCACGAGCTTTATGAAC 779
286 rAsnSerThr.....ileGlnGluSerLeuHisArgSerG 298
      : : : : : ||| : : : : : ||| : : : : :
780 GAACACAAGTGCAGGCAATCGATATCTATAAAAAAAGTGTACACAGTG 829
298 lu...CysProSerGlnArgTyIleSerLeuGlnCysSerHisCysGly 313
      : : : : : ||| : : : : : ||| : : : : :
830 ATGCCTGTTCTTCAAAAGCAGTGGTTCTTTACGCTGTATAGCTGCGGG 879
314 LeuArg.....AlaMetThrGlyArgIleValGlyAlaLeuAl 327
      : : : : : ||| : : : : : ||| : : : : :
880 GTCAACTTGAACTCAGCCCGCAGAGAGGATGTGGCGCGCAGAGCGC 929
327 aSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThrH 344
      : : : : : ||| : : : : : ||| : : : : :
930 GCTCCCGGGGCGCTGGCCCTGGCAGGTGCAGCTGCACGTCAGAACGTC 979
344 isoleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAla 360
      : : : : : ||| : : : : : ||| : : : : :
980 ACGTGTGCGGAGGCTCCATCATCCCCCGAGTGGATCGTGACAGCGCC 1029
361 HisCysPhePheValThrArgGluLysValLeuGluGly.....Tr 374
      : : : : : ||| : : : : : ||| : : : : :
1030 CACTGCGTG.....GAAACCTCTTAAACAATCCATGGCATTG 1067
374 pLysValTyAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAla. 390
      : : : : : ||| : : : : : ||| : : : : :
1068 GACGGCATTTGCGGGATTTTGAGACAAATCTTTCATGTTCTATGGAGCG 1117
391 .....SerIleAlaGluIleIleAsnSerAsnTyThrAspGluGlu 405
      : : : : : ||| : : : : : ||| : : : : :
1118 GATACCAAGTAGAAAGTAGTGTCTTCAATCAATATGACTCCAGACC 1167
406 AspAspTyAspIleAlaLeuMetArgLeuSerLysProLeuThrLeuSe 422
      : : : : : ||| : : : : : ||| : : : : :
1168 AAGACAATGACATTCGCTGATGAAGTGCAGAGCCCTGTGACTTTCAA 1217
422 rAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThrPheSerL 439
      : : : : : ||| : : : : : ||| : : : : :
1218 CGACCTAGTGAACACCAAGTGTGTGCCACCCAGGATGATGCTGCAGC 1267
439 euAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAsp 455
      : : : : : ||| : : : : : ||| : : : : :
1268 CAGACAGCTCTCTGATTTCCGGTGGGGGCCACCGAGGAG...AAA 1314
456 AspLysThrSerProPheLeuArgGluValGlnValAsnLeuIleAsp 472
      : : : : : ||| : : : : : ||| : : : : :
1315 GGGAAGACCTCAGAAAGTGTGAACGTCGCCAAGTGTCTTCATTGAGAC 1364
472 eLysLysCysAsnAspTyLeuValTyLeuAspSerTyLeuThrProArgM 489
      : : : : : ||| : : : : : ||| : : : : :
1365 ACAGAGATGCAACAGCAGATATCTCTATGACAACTGATCACACCAACCA 1414
489 etMetCysAlaGlyAspLeuArgGlyArgAspSerCysGlnGlyAsp 505
      : : : : : ||| : : : : : ||| : : : : :
1415 TGATCTGCGCGGCTCTTCGAGGAGGAGCGTCGATCTTCCCGAGGGTAC 1464
506 SerGlyGlyProLeuValCysGluGlnAsnAsnArgTrpTyLeuAlaG 522
      : : : : : ||| : : : : : ||| : : : : :
1465 AGTGGAGGCGCTCTGCTCACTTCGAGAACAAATATCTGGTGGCTGATAG 1514
522 yValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyVal 539
      : : : : : ||| : : : : : ||| : : : : :
1515 GGATACAAGCTGGGGTCTGCTGTGCCAAAGCTTACAGACCAAGGAGTGT 1564
539 yrThrLysValThrGluValLeuProTrpIleTySerLysMetGluSer 555
      : : : : : ||| : : : : : ||| : : : : :
1565 ACGGGAATGTGATGTTATTCACGGACTGGATTTATCGACAAATGAGGGCA 1614
```

```
556 Glu 556
      : : :
1615 GAC 1617
```

seq_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH93942

seq_documentation_block:

ID AAH93942 standard; cDNA; 3245 BP.

XX AAH93942;

XX 04-OCT-2001 (first entry)

XX P1000C full length cDNA sequence.

XX Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis; ss.

XX Homo sapiens.

XX WO200151633-A2.

XX 19-JUL-2001.

XX 16-JAN-2001; 2001WO-US01574.

XX 14-JAN-2000; 2000US-0483672.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jfang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;

XX WPI; 2001-425873/45.

XX New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines.

PS Claim 1; Page 537-538; 543pp; English.

XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used to treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAH01115 to
CC AAH01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.

XX SQ Sequence 3245 BP; 774 A; 818 C; 870 G; 783 T; 0 other;

alignment_scores:

Quality: 743.00 Length: 551
Ratio: 2.173 Gaps: 17
Percent Similarity: 62.069 Percent Identity: 33.394

alignment_block:

US-09-879-792-12 x AAH93942 ..

Align seg 1/1 to: AAH93942 from: 1 to: 3245

43 GlnAlaSerProAlaGlyThrProGlyArgAlaSerProAlaGlnAl 59
: : : : : ||| : : : : :
: : : : : ||| : : : : :


```

282 ..... 282
152 hehrrtrpArgGluGlyGlnLysGlnLeuProLeuIleGlyCysValLeu 168
283 .....CTCGTGGAGCT..... 294
169 LeuLeuIleAlaLeuValValSerLeuIleIleLeuPheGlnPheTrpG1 185
295 .....GCGCTGGCGCGTGGCCCTACTC.....TGGAA 320
185 n.....GlyHisThrGlyIleArgTyrLysGluGlnA 196
321 GTTCATGGCAGCAAGTGTCCAACTCTGGGATAGAGTGC...GACTCCT 367
196 rgGluSerCysProLysHisAlaValArgCysAspGlyValValAspCys 212
368 CAGGTACTTGATCAACCCCTCTAACTGGTGTGATGGCGTGTCACTGC 417
213 LysLeuLysSerAspGluLeuGlyCysValArgPheAspTrpAspLys 229
418 CCGCGCGGGAGGACGAGAACTCGTGTTCGCTCTACGGATCAAACTT 467
229 rLeuLeuLysIleTyrSerCysSerHisGlnTrpLeuProIleCys 246
468 CAPCCTTCAGGTGTACTCTCATCTCAGAGGAAGTCTTGGCCCTGTGTGCC 517
246 erSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGlnLeuGly 262
518 AAGACGACTGGAACGAGAACTACGGCGGGCGGCTGACGGGACATGGC 567
263 PheGluSerAlaHisArgThrThrGluValAlaHisArgAspPheAla 279
568 TATAAGAAATAATTTTACTCTAGCCAA...GGAATAGTGGATGACAGCGG 614
279 nSerPheSerIleLeuArgTyrAsnSerThr.....IleG 291
615 ATCCACGAGCTTATGAAACTGAACACAGTCCGCGCAATGTGATATCT 664
291 lngLusSerLeuHisArgSerGlu...CysProSerGlnArgTyrIleSer 306
665 ATAAAAACTGTACCACAGTATGATGCTCTTCAAAAGCAGTGGTTCT 714
307 LeuGlnCysSerHisCysGlyLeuArg.....AlaMetThrGlyAr 320
715 TTACGCTGTATACCTTCGCGGGTCAACTTCAACTCAAGCGCCAGAGCAG 764
320 gIleValGlyGlyAlaLeuAlaSerAspSerLysTyrProTrpGlnVal 337
765 GATGTGGCGCGGAGAGCGCGCTCCCGGGGGCTGGCCCTGGCAGGTCA 814
337 erLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuIleAspAla 353
815 GCCTGCAGCTCCAGACAGCTCCAGTGTGGAGGCTCCATCATCACCCC 864
354 GlnTrpValLeuThrAlaAlaHisCysPhePheValThrArgGluLysVa 370
865 GACTGGATCGTGACAGCGGCCACTCGGTG.....GAAAAACC 902
370 lLeuGluGly.....TrpLysValTyrAlaGlyThrSerAsnLeuH 384
903 TCTTAACAATCCATGCATTTGGAGCGCATTTGCGGGGATTTTGAGACAAT 952
384 isGlnLeuProGluAlaAla.....SerIleAlaGluIleIleIleAsn 398
953 CTTTCATGTTCTATGAGCGGGATACACAGTGAAGAAAGTATCTTCAT 1002
399 SerAsnTyrThrAspGluAspAspTyrAspIleAlaLeuMetArgLe 415
1003 CCAAAATATGACTCCAGACCAAGACATGACATTTGCGCTGATGAGCT 1052
415 uSerLysProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProM 432

```

```

1053 GCAGAGCCCTCTGACTTTCAACGACCTAGTGAAACACAGTGTGTGCCCA 1102
432 ethisGlyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPhe 448
1103 ACCCAGCATGATGCTGCCACGACAGCTCTGCTGGATTTCGGGTGG 1152
449 GlyLysThrArgGluThrAspAspLysThrSerProPheLeuArgGluVa 465
1153 GGGCCACCCAGGAG...AAAGGGAAGACCTCAGAACTGTGAACGCTGC 1199
465 lGlnValAsnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrA 482
1200 CAAGGTCTTTCATTGACACAGAGATGCAACAGCAGATATGCTCATG 1249
482 spSerTyrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyGly 498
1250 ACAACCTGATCACACACGATGATCTGTCCGCTTCTCCTGCAGGGAAC 1299
499 ArgAspSerCysGlnGlyAspSerGlyGlyProLeuValCysGluGlnAs 515
1300 GTCGATTTCTTCCAGGGTGACAGTGGAGGCGCTCTGGTCTCAGAAAGAA 1349
515 nAsnArgTyrTyrLeuAlaGlyValThrSerTyrGlyThrGlyCysGlyG 532
1350 CAATATCTGGTGTGATAGGGATACAACTGGGTCTTGGCTGTGCCA 1399
532 lnArgAsnLysProGlyValTyrThrLysValThrGluValLeuProTrp 548
1400 AAGCTTACAGCACGAGAGTGTACGGGAATGTGATGGTATTTCACGGACTGG 1449
549 IleTyrSerLysMetGluSerGlu 556
1450 ATTTATCGACAAATGAGGCGAGAC 1473

```

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAS64179

seq_documentation_block:

ID AAS64179 standard; cDNA; 1479 BP.

```

XX AAS64179;
AC
DT 29-JAN-2002 (first entry)
XX
DE Human prostate cDNA sequence #598.
XX
KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedwick TS, Carter D;
PI Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.

```


[illegible]

OM of: US-09-879-792-12 to: EST:* out_format : pfs

Date: Aug 20, 2002 4:54 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-O=/cgn2_1/USPTO_pool/US09879792/tunat_15082002_104127_22225/app_query.fasta_1.629
-DB=EST -ORFMT=FASTAP -SUFFIX=rst -GAPOP=12,000 -GAPEXT=4,000
-MINWATCH=0,100 -LOOPL=0,000 -LOOPEXT=0,000 -QGAPOP=4,500
-QGAPEXT=0,050 -XGAPOP=10,000 -XGAPEXT=0,500 -FGAPOP=6,000
-FGAPEXT=7,000 -YGAPOP=10,000 -YGAPEXT=0,500 -DELOP=6,000
-DELETEXT=7,000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09879792_@CENL1_1_5172
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-879-792-12
Query length: 562
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1888.370000
```

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_est2:BE732381	+	1282.50	1028.78	4.7e-48	834	1 BE732381 601569601F1 NIH_MGC_21
gb_est2:BE289529	+	893.00	721.44	6.1e-31	678	1 BE289529 601080601F1 NCI_CGAP_M
gb_est2:BE286322	+	845.50	684.53	7.0e-29	614	1 BE286322 601095164F1 NCI_CGAP_M
gb_est2:BF119003	+	749.00	607.89	1.3e-24	622	1 BF119003 601755367F1 NCI_CGAP_M
gb_est2:BE285038	+	708.50	575.78	8.0e-23	621	1 BE285038 601098725F1 NCI_CGAP_M
gb_est2:R78581	+	626.00	513.73	2.3e-19	402	1 R78581 Y473c10.r1 Soares placen
gb_est2:BE290088	+	615.00	502.37	9.8e-19	563	1 BE290088 601080566F1 NCI_CGAP_M
gb_hlc:AK004939	+	592.00	471.01	5.5e-17	3030	1 AK004939 Mus musculus adult ma
gb_est2:BE280394	+	573.00	474.74	3.4e-17	751	1 BE280394 601158674F1 NIH_MGC_21
gb_hlc:AK002694	+	573.00	463.11	1.5e-16	1814	1 AK002694 Mus musculus adult ma
gb_hlc:BC004855	+	573.50	459.68	2.3e-16	1973	1 BC004855 Homo sapiens, Similar
gb_hlc:BC008514	+	555.00	441.12	2.5e-15	3248	1 BC008514 Mus musculus, suppress
gb_hlc:AK005546	+	541.50	433.29	6.9e-15	2246	1 AK005546 Mus musculus adult fe
gb_est2:BE309103	+	529.00	432.59	7.5e-15	688	1 BE309103 601093331F1 NCI_CGAP_M
gb_est1:AL555870	+	513.50	418.42	4.6e-14	876	1 AL555870 AL555870 LTI_NFL006.PL
gb_est1:AW845106	+	508.00	422.81	2.6e-14	285	1 AW845106 MRO-CT0006-010699-022
gb_est1:AL578261	+	507.50	413.14	9.1e-14	936	1 AL578261 AL578261 LTI_NFL006.PL
gb_hlc:AK006271	+	505.00	410.26	1.3e-13	1050	1 AK006271 Mus musculus adult ma
gb_est1:BI771604	+	497.50	407.80	1.8e-13	671	1 BI771604 603058729F1 NIH_MGC_12
gb_est1:AI909842	+	484.00	403.74	3.1e-13	286	1 AI909842 QV-BT225-050599-056 BT
gb_est1:AL551470	+	481.00	391.61	1.4e-12	999	1 AL551470 AL551470 LTI_NFL006.PL
gb_est2:BI554641	+	480.50	390.81	1.6e-12	1052	1 BI554641 603233988F1 NCI_CGAP_M
gb_est2:BI259237	+	479.50	392.38	1.3e-12	777	1 BI259237 602970095F1 NIH_MGC_12
gb_est2:BG288427	+	473.50	385.85	3.0e-12	975	1 BG288427 602388091F1 NIH_MGC_93
gb_hlc:AK016509	+	471.50	383.40	4.0e-12	1089	1 AK016509 Mus musculus adult ma
gb_est2:BI220817	+	457.50	376.33	1.0e-11	649	1 BI220817 602937335F1 NCI_CGAP_M
gb_hlc:AK010640	+	454.50	366.78	3.5e-11	1629	1 AK010640 Mus musculus ES cells
gb_est2:BI433250	+	452.50	369.00	2.6e-11	1000	1 BI433250 602986390F1 NCI_CGAP_M
gb_hlc:AK007987	+	443.00	355.78	1.4e-10	2073	1 AK007987 Mus musculus 10 day c
gb_est1:AW202972	+	439.50	361.03	7.3e-11	740	1 AW202972 f24q04.y1 Zebrafish a
gb_est2:BF789214	+	433.50	353.88	1.8e-10	1006	1 BF789214 602105109F1 NCI_CGAP_M
gb_est2:BF144893	+	431.50	353.50	1.9e-10	862	1 BF144893 602909736F1 NCI_CGAP_M
gb_est2:BF757612	+	430.00	361.90	6.5e-11	252	1 BF757612 IL0-CT3002-071100-500
gb_est2:BF757611	+	430.00	361.83	6.6e-11	254	1 BF757611 IL0-CT3002-071100-500
gb_est2:BF757610	+	430.00	361.80	6.6e-11	255	1 BF757610 IL0-CT3002-071100-500
gb_est2:BG915045	+	427.50	349.75	3.1e-10	928	1 BG915045 602814325F1 NIH_CGAP_M
gb_est2:BI827608	+	426.00	348.99	3.4e-10	878	1 BI827608 603073575F1 NIH_MGC_11
gb_est2:BI334059	+	425.50	348.66	3.6e-10	871	1 BI334059 602997320F1 NIH_MGC_12
gb_est1:AK014645	+	425.50	345.40	5.4e-10	1323	1 AK014645 Mus musculus 10 days
gb_est2:BG679249	+	424.50	348.49	3.6e-10	804	1 BG679249 602627332F1 NCI_CGAP_S

```
gb_est2:BG754441 + 421.00 344.69 5.9e-10 916 1 BG754441 602710073F1 NIH_MGC
gb_est1:BE200991 + 419.50 347.52 4.1e-10 547 1 BE200991 fk84e06.y1 Zebrafis
gb_est2:BM487066 + 419.50 346.64 4.6e-10 613 1 BM487066 pgm2n.pk003.h18 Nor
gb_est2:BG966811 + 419.00 345.33 5.5e-10 689 1 BG966811 602834306F1 NCI_CGA
gb_hlc:AK006695 + 419.00 339.24 1.2e-09 1504 1 AK006695 Mus musculus adult

seq_name: gb_est2:BE732381

seq_documentation_block:
LOCUS BE732381 834 bp mRNA linear EST 15-SEP-2000
DEFINITION 601569601F1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3844129 5',
mRNA sequence.
ACCESSION BE732381
VERSION BE732381.1 GI:10146373
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC540 row: b column: 02
High quality sequence stop: 817.
FEATURES
Location/Qualifiers
1..834
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3844129"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life technologies)."
```

```
BASE COUNT 186 a 246 c 224 g 178 t
ORIGIN
alignment_scores:
Quality: 1282.50 Length: 282
Ratio: 4.750 Gaps: 4
Percent Similarity: 95.745 Percent Identity: 93.617
alignment_block:
US-09-879-792-12 x BE732381 ..
Align seg 1/1 to: BE732381 from: 1 to: 834
219 LeuGlyCysValArgPheAspTrpAspLysSerLeuLeuTyrSe 235
|||||
2 CTGGCGTCGGGAGGTGTGACTGGGACAAAGTCTCTGCTTAAATCTACTC 51
|||||
235 rGlySerSerHisGlnTrpLeuPheProIleCysSerSerAsnTrpAsnAsps 252
|||||
52 TGGGTCCTCCCATCAGTCGGCTTCCCATCTGTAGCAGCACTGGAATGACT 101
|||||
252 eTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArg 268
|||||
```

```
102 CCTACTCAGAGACAGCTGCCAGCAGCTGGGTTTCGAGAGTGCTCACCGG 151
269 ThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuAr 285
|||||
152 ACAACCCAGGTGTCACACAGGATTTTGCCAAACAGCTTCAATCTTGAG 201
285 gTyrAsnSerThrIleGlnGluSerLeuHisArgSerGluCysProSerG 302
|||||
202 ATACAACCTCCACCATCCAGGAAGCCCTCCACAGGTCTGAATGCCCTTCCC 251
302 InArgTyrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMetThr 318
|||||
252 AGCGGTATATCTCCTCCATGTTCCCACTGCGGACTGAGGGCCATGACC 301
319 GlyArgIleValGlyGlyAlaLeuAlaSerAspSerLysTrpProTrpG1 335
|||||
302 GGGCGGATCTGGGAGGGCGCTGGCCCTCGGATAGCAAGTGGCCTTGCCA 351
335 nValSerLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuIleA 352
|||||
352 AGTGAGTCTGCACTTCGGCACCAACCCACATCTGTGGAGGACGCTCATTG 401
352 spAlaGlnTrpValLeuThrAlaAlaHisCysPhePheValThrArgGlu 368
|||||
402 ACGCCCAAGTGGGTGCTCACGCGCCGCCACTGCTTCTTCGTGAGCCGGGAG 451
369 LysValLeuGluGlyTrpLysValTyrAlaGlyThrSerAsnLeuHisG1 385
|||||
452 AAGGTCCTGGAGGGCTGGAAGGTGTACGCGGGCCACCAAGCAACTG.CACCA 500
385 nLeuProGluAlaAlaSerIleAlaGluIleIleAlaAsnSerAsnTyrT 402
|||||
501 GTTGCTCTGAGGCAGC.TCCATTGCCGAGATCATCATCAACAGCAATTACA 549
402 hrAspGluGluAspAspTyr.AspIleAlaLeuMetArgLeu.SerLysP 418
|||||
550 CCGATGAGGAGGAGCAGCATTTGACATCGCCCTCATGCGGCTGTTCCAAGC 599
418 roLeuThrLeuSerAlaHisIleHisProAlaCysLeu.ProMetHisG1 434
|||||
600 CCCTGAACCTGTCCGT.CACATCCACCCTGCTTGCCCTCCCCCATGCATGG 648
434 yGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysT 451
|||||
649 ACAGACCTTTAGCTCCTCAATGAGACCTCTTGATCACAGGCTTTTGGCAAG 698
451 hrArgGluThrAspAspLysThrSerProPheLeuArgGluValGlnVal 467
|||||
699 ACAGGAGACAGATGAANAACACATCCCCCTTCCT.CGGGAGGTGCAGGTC 747
468 AsnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrAspSerty 484
|||||
748 AATCTCATGACTTCCAGAAATGCAATGACTTAACGTGGTCTATGACAGTAC 797
484 rLeuThrProArgMetMetCysAlaGlyAspLeuArgGly 497
|||
798 CTTACCAAGAGATGATGTGT...GGGGAACCTTCGTGGG 834
```

seq_name: gb_est2:BE289529

```
seq_documentation_block:
LOCUS BE289529 678 bp mRNA linear EST 26-OCT-2000
DEFINITION 601088060F1 NCI_CGAP_Mam5 Mus musculus cdna clone IMAGE:3482919 5',
mRNA sequence.
ACCESSION BE289529
VERSION BE289529
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 678)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cdna Library Preparation: Life technologies, Inc.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8513 row: g column: 16
High quality sequence stop: 615.

FEATURES

source

1..678
Location/Qualifiers
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 148 a 207 c 183 g 140 t
ORIGIN

alignment_scores:

Quality: 893.00 Length: 208
Ratio: 4.533 Gaps: 0
Percent Similarity: 94.712 Percent Identity: 78.846

alignment_block:

US-09-879-792-12 x BE289529 ..

Align seg 1/1 to: BE289529 from: 1 to: 678

```
104 SerAlaSerValThrThrSerProThrArgValTyrLeuValArgAlaTh 120
|||||
1 TCAGCCTCCAGCATCTCCCCAACGAGAGTGATCCTTGTAGAGCAAC 50
120 rProValGlyValaValProIleArgSerSerProAlaArgSerAlaProA 137
|||||
51 ACCAGTGGGGCTGTCCCATCCGGGCATCTCTGCCAGGTCACACCAG 100
137 laThrArgAlaThrArgGluSerProGlyThrSerLeuProLysPheThr 153
|||||
101 CCACCCAGGGCCACCAGGGAGAGCCAGGCTCTCAGTTTCCCAAGTTCTCC 150
154 TrpArgGluGlyGlnLysGlnLeuProIleGlyCysValLeuLeuLe 170
|||||
151 TGGCAGGAGAGCCAGAGGAGCTGCCACTCATCGGGTGTGCATCCTTCT 200
170 uIleAlaLeuValValSerLeuIleIleLeuPheGlnPheTrpGlnGlyH 187
|||||
201 CATCAGCCTGGTGATCTCGCTCATCTCTTCTTACTTCTGGAGAGGCC 250
187 isThrGlyIleArgTyrTlyGluGlnArgGluSerCysProLysHisAla 203
|||||
251 ACACAGGGATCAAGTACAAAGAGCCACTGGAGAGTTGCCCTATCCAGCCA 300
204 ValArgCysAspGlyValValAspCysLysLysSerAspGluLeuG1 220
|||||
301 GTTCGCTGTGATGGAGTGGGACTGCAAAATGAAGACCGATGAGCTGGG 350
220 yCysValArgPheAspTrpAspLysSerLeuLeuLysIleTyrSerGlyS 237
|||||
351 CTGTGTCAAGTTCGATGGGCAAGTCCCTCCTGAAAGTCTACTCTGGGT 400
```

```
237 erSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyr 253
||||| ::::::::::::::::::::::::::::::::::::::::::::
401 CTTCTGGGAGTGCTTCTCTGTCAGCAGCAGCTGGAACGACACTGAC 450

254 SerGluLysThrCysGlnGlnLeuGluPheGluSerAlaHisArgThrTh 270
||||| ::::::::::::::::::::::::::::::::::::::::::::::
451 TCCAAGAGGACCTGCCACGAGCTGGGATTTGACAGCGCTTACCGAACAAC 500

270 rGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrA 287
||||| ::::::::::::::::::::::::::::::::::::::::::::::
501 TGAGGTAGCCACAGGACATCACACGACGCTTCTTACTTCCGAATACA 550

287 snSerThrIleGlnGlnSerLeuHisArgSerGluCysProSerGlnArg 303
||||| ::::::::::::::::::::::::::::::::::::::::::::::
551 ACACCACCATCCAGGAAGGCTCTACAGGTGCGCAATGCTCTCCGG.CGG 599

304 TyrIleSerLeuGlnCysSerHis 311
||||| ::::::::::::::::::::::::::::::::::::::::::::::
600 TAGTCTCCCTCCAGTGTGTTCCAC 623
```

seq_name: gb_est2:BE286322

```
seq_documentation_block:
LOCUS BE286322 614 bp mRNA linear EST 26-OCT-2000
DEFINITION 601095164F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490022 5',
mRNA sequence.
ACCESSION BE286322
VERSION BE286322.1 GI:9164563
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 614)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM8531 row: 0 column: 15
High quality sequence stop: 569.
Location/Qualifiers
1..614
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3490022"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
```

FEATURES

source

seq_name: gb_est2:BF119003

```
seq_documentation_block:
LOCUS BF119003 622 bp mRNA linear EST 24-OCT-2000
DEFINITION 601755367F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3984761 5',
mRNA sequence.
ACCESSION BF119003
VERSION BF119003.1 GI:10958043
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 622)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
```

BASE COUNT 159 a 168 c 161 g 126 t
ORIGIN

alignment_scores:

Quality: 845.50 Length: 170
Ratio: 5.093 Gaps: 2
Percent Similarity: 97.647 Percent Identity: 91.765

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9187 row: e column: 18

High quality sequence stop: 618.

FEATURES

Source

1. .622

/organism="Mus musculus"

/strain="C57/B6"

/db_xref="taxon:10090"

/clone="IMAGE:3984761"

/clone_lib="NCI_CGAP_Mam5"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH"

BASE COUNT 135 a 147 c 185 g 155 t

ORIGIN

alignment_scores:

Quality: 749.00 Length: 137

Ratio: 5.507 Gaps: 0

Percent Similarity: 99.270 Percent Identity: 96.350

alignment_block:

US-09-879-792-12 x BF119003 ..

Align seg 1/1 to: BF119003 from: 1 to: 622

426 HisProAlaCysLeuPromethHisGlyGlnThrPheSerLeuAnsluTh 442

2 CACCCTGCTCCCTCCCATCATGGTCAGACCTTGGCCCTCAATGAGAC 51

442 rCysTrpIleThrGlyPheGlyLysThrArgGluThrAspLysThrs 459

52 CTGTTGGATCACGGGCTTTGGCAAAACCAAGGACAGATGAGAGACAT 101

459 erProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysCys 475

102 CTCCCTTCCTCCGAGAGGTTCCAGGTCAACCTCATGTTCACAGAGTGC 151

476 AsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAl 492

152 AATGACTACTTGGTCTATGACAGCTACCTTACCCCAAGGATGATGTGTC 201

492 aGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyP 509

202 CGGGGATCTACGAGGAGGGAGGACTCTCCGAGGGAGACAGTGGAGGAC 251

509 roLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSer 525

252 CTCCTCGTCTGTGACGAGAACAACTCGCTGGTACCTGGCAGGTGTACCA 301

526 TrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysVa 542

302 TGGGACAGGCTGTGGCCAGAAAAACAAGCCTGGTGTGTACACAAAGT 351

542 lThrGluValLeuProTrpIleTyrSerLysMetcluserGluValArgp 559

352 GACAGAAGTACTTCCCTGGATTTACAGAAAGATGGAGAGTGGAGTACCAT 401

559 heArgLysSer 562

402 TCCGGAATCT 412

seq_name: gb_est2:BE285038

seq_documentation_block:

LOCUS BE285038

621 bp mRNA linear

EST 26-OCT-2000

DEFINITION 601098725F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3497358 5', mRNA sequence.

ACCESSION BE285038

VERSION BE285038.1 GI:9162286

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 621)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8551 row: a column: 07
High quality sequence stop: 529.

FEATURES

source

1. .621

Location/Qualifiers

/organism="Mus musculus"

/strain="C57/B6"

/db_xref="taxon:10090"

/clone="IMAGE:3497358"

/clone_lib="NCI_CGAP_Mam5"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH"

BASE COUNT 146 a 184 c 163 g 128 t

ORIGIN

alignment_scores:

Quality: 708.50 Length: 207

Ratio: 3.851 Gaps: 8

Percent Similarity: 88.889 Percent Identity: 74.396

alignment_block:

US-09-879-792-12 x BE285038 ..

Align seg 1/1 to: BE285038 from: 1 to: 621

98 ArgSerSerSerAlaArgSerAlaSerValThrThrSerProThrArgVa 114

1 AGATCATCATCTGCCAGGTCCAGCTCCACGACATCTCCCAACGAGAGT 50

114 lTyrLeuValArgAlaThrProValGlyAlaValProIleArgSerSerP 131

51 GTACCTTTAGAGCACACACAGTGGGGCTGTCCCATCCGGGCATCTC 100

131 roAlaArgSerAlaProAlaThrArgAlaThrArgGlu, SerProGlyTh 147

101 CTGCCAGGTTCAGCACCCAGCCAGGCGCCACCGAGGTAGCCAGGCTCT 150

147 rSerLeuProLysPheThrTrp, ArgGluGlyGlnLysGlnLeuProLeu 163

151 CAGTTTCCCAAGTTCTCTCTGGTCAGGAGACCCAGAGGCGAGCTGCAC 200

164 lIleGlyCysValLeuLeuLeuIleAlaLeuValValSerLeuIleIlele 180

201 ATCGGGTGTGTCATCTTCTCATCAGCCTGGTGTGATCTCGCTCATCTTCT 250


```

180 uPheGlnPheTrp.GlnGlyHisThrGlyIleArgTyrLysGluGlnArg 196
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
251 CTTTACTTCTGGAGAGTGGCCACAGAGGATCAAGTACAAAGAGCCACTG 300
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
197 GluSerCysProLysHisAlaValArgCysAspGlyValValAsp.CysL 213
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
301 GAGAGTTCCTTATCCACGAGTTCGCTGTGATGGAGTGGTGGACTTGC 350
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 ySLeuLysSerAspGluLeuGlyCysValArgPheAspTrpAspLysSer 229
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
351 AAATGAAGAGCGATGAGCTGGCTGTCTCAGGTTTCGACTGGCAAGTCC 400
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
230 LeuLeuLysIleTyrSerGlySerSer.HisGlnTrpLeuProIleCys 246
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
401 CTCCTGAAGTCTACTCTGGGTCTCTCTGGCAGAGTGGCTTCTCTGTCA 450
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
246 erSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGln.LeuG 262
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
451 GCAGCAGCGG.AACGACACTGACTCCAGAGGAGCTGCCAGCAAGCTGGG 499
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
262 yPheGluSerAlaHisArgThrThrGluValAlaHisArgAspPheAla 279
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
500 ATTGACAGCGCTTACCGAACACTGAGGTAGCCACAGACATCACCA 549
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
279 snSerPhe.SerIleLeuArgTyrAsnSerThrIleGlnGluSerLeuH 295
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
550 GCAGCTTCTAACTCTCGGAACAAACAAAC...ATCCAGGAAGGCTCTA 596
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
295 sArgSer 297
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
597 CAGTCG 603

```

seq_name: gb_est2:R78581

seq_documentation_block:

LOCUS R78581 402 bp mRNA linear EST 09-JUN-1995
DEFINITION Y173c10.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:144882 3' similar to SP:HEPS_RAT Q05511 SERINE PROTEASE
HEPSIN ; contains MER22 repetitive element ; , mRNA sequence.

ACCESSION R78581

VERSION R78581.1 GI:854862

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 402)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 554

High quality sequence stops: 312

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Insert Length: 554

Seq primer: Promega -21m13

High quality sequence stop: 312.

FEATURES

Location/Qualifiers

1..402

/organism="Homo sapiens"

/db_xref="GDB:558601"

/db_xref="taxon:9606"
/clone="IMAGE:144882"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AACTGGAAGAATTCCGGCGCGAGGAATTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 76 a 124 c 117 g 82 t 3 others
ORIGIN

alignment_scores:

Quality: 626.00 Length: 125
Ratio: 5.131 Gaps: 2
Percent Similarity: 97.600 Percent Identity: 96.800

alignment_block:

US-09-879-792-12 x R78581 ..

Align seg 1/1 to: R78581 from: 1 to: 402

```

299 CysProSerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLeuAr 315
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3 TGCCCTTCCCGAGCGGTATATCTCCCTCCAGTGTTCCTCCACTCGGACTGAG 52
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
315 gAlaMetThrGlyArgIleValGlyGlyAlaLeuAlaSerAspSerLysT 332
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
53 GGCCATGACCGGGCGGATCGTGGAGGGCGCTGGCTCGGATAGCAAGT 102
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
332 rpProTrpGlnValSerLeuHisPheGlyThrHisIleCysGlyGly 348
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
103 GGCCTTGGCAAGTGGCTGCACTTTCGGCACCACCCACATCTGTGGAGGC 152
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
349 ThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPheVa 365
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
153 ACGCTCATTTGACGCCAGCTGGTGTCTCATTNCCGCCCTCTCTTCGT 202
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
365 1.ThrArgGluLysValLeuGluGlyTrpLysValTyrAlaGlyThrSer 381
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
203 GNACCCGGGAGAAAGTCTGGAGGGCTGGAAAGGTACGCGGGACCAGC 252
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
382 AsnLeuHisGlnLeuProGluAlaAlaSerIleAlaGluIleIleAs 398
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
253 AACCTGCACCAAGTTCCTTGAGGAGCCCTCCATTGCCGAGATCATCAA 302
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
398 nSerAsnTrpThrAspGluGluAspTyrAspIleAlaLeuMetArgL 415
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
303 CAGCAATTATACCGATGAGGAGGAGCAGCTATGACATCGCCCTCATGCGGC 352
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
415 eu.SerLysProLeuThrLeu 421
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
353 TGTTCAGNCCCTTGACCCCTG 373

```

seq_name: gb_est2:BE290088

seq_documentation_block:

LOCUS BE290088 563 bp mRNA linear EST 26-OCT-2000
DEFINITION 601089056F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3483900 5',
mRNA sequence.

ACCESSION BE290088

VERSION BE290088.1 GI:9170901

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 563)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM8515 row: p column: 13
 High quality sequence stop: 515.

FEATURES

source
 1. .563
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:3483900"
 /clone_lib="NCL_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 113 a 191 c 147 g 112 t
 ORIGIN

alignment_scores:
 Quality: 615.00 Length: 192
 Ratio: 3.796 Gaps: 10
 Percent similarity: 84.375 Percent Identity: 71.354
 alignment_block:
 US-09-879-792-12 x BE290088 ..
 Align seg 1/1 to: BE290088 from: 1 to: 563

81 ProAlaArgAlaSerProAlaLeuAlaSerLeuSerArgSerSerGl 97
 |||
 2 CCAGCCGGGCATCTCCAGCCAGGCACCACCTCCAGGTCATCATCGGC 51
 97 yArgSerSerAlaArgSerAlaSerValThrThrSerProThr.Arg 113
 : |||
 52 AGCATCATCATCTGCCAGGTCAGGCTCCAGGACATCTCCCAACGCAGA 101
 114 ValTyrLeuValArgAlaThrPro...ValGlyAlaValProIleArgSe 129
 |||
 102 GTGTACCTGTTAGAGCAACACAGCTGCGGGTGCTGTCCCATCCGGGC 151
 129 rSerProAlaArgSerAlaProAlaThrArgAlaThrArgGlu.SerPro 145
 : |||
 152 ATCTCTGCCAGGTCAGCACGCCACCGAGGCCACCGAGGAGCGGCCA 201
 146 GlyThrSerLeu.ProLysPheThrTrpArgGluGlyGlnLysGlnLeuP 162
 |||
 202 GGTCATCAGTCTCCCAAGTCTCTCTGGCAGGAGACCAGAGCGAGCTGC 251
 162 roLeuIleGlyCysValLeuLeuLeuIleAlaLeuValSerLeuIle 178
 |||
 252 CACTCAFCGGGTGTGTCATCTCTTCATCAGCCTGGTGATCTCGTCATC 301
 179 IleLeuPheGlnPheTrpGlnGlyHisThr.GlyIleArgTyrLysGluG 195
 : |||
 302 CTCTCTCTTACTTCTGGAGAGGCCACACAGGGGCATCAAGTACAAAGAGC 351

195 lnArgGluSerCysProLysHisAlaValArgCysAspGly.ValValAs 211
 |||
 352 CACTGGAGAGTTGCCCTATCCACGAGTTCGCTGTGATGGAGCTGGTGA 401
 211 pCysLysLeuLys.SerAspGlnLeu.GlyCysValArg...PheAspTr 226
 |||
 402 CTGCAAAATGAACGAGCGATGAGCTGGAGTGTGTCAACGGCTTCGACTG 451
 226 pAspLysSerLeuLeu.LysIleTyrSerGlySerSerHisGlnTrpLeu 242
 |||
 452 GGACAAGTCCCTCTCCAGCAGCTACTCTGGTCTTCTGGCGAGTGGCTT 501
 243 ProIleCysSerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGl 259
 |||
 502 CCTCTCTGCAGCAGCTGGAGACGACACTGACTCCACGAGGACTGCCA 551

259 nGlnLeuGly 262
 |||
 552 GCAGCTGGGC 561

seq_name: gb_htc:AK004939

seq documentation_block:

LOCUS AK004939 3030 bp mRNA linear HTC 19-JAN-2002
 DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
 library, clone:1300008A22:homolog to DJ1170K4.4 (NOVEL PROTEIN)
 (FRAGMENT), full insert sequence.

ACCESSION AK004939

VERSION AK004939.1 GI:12836502

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
 clone_lib:RIKEN full-length enriched mouse cDNA library
 clone:1300008A22.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (sites)
 Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE

2 (sites)
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE

3 (sites)
 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE

4 (sites)
 The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5 (bases 1 to 3030)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
 Arakawa,T., Balderelli,R., Bono,H., Brownstein,M., Bult,C.,
 Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
 Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,

Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koyra,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tegan,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE
JOURNAL
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cdna library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cdna was primed with a primer [5' GAGAGAGAGCGCGCACTCGAGTGTGTTTTTTTTTNN 3'], cdna was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cdna was prepared with the primer adapter of sequence[5' GAGAGAGAGAGTTCAGAGCTCAATTATTAATTAACCCCCCCC 3']. cdna was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

Location/Qualifiers
 1..3030
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:MG1:1904397"
 /db_xref="taxon:10090"
 /clone="1300008A22"
 /sex="male"
 /tissue.type="liver"
 /clone_lib="RIKEN full-length enriched mouse cdna library"
 /dev_stage="adult"
 91..2490
 /note="data source:SPTR, source key:095519, evidence:ISS homolog to DJ1170K4.4 (NOVEL PROTEIN) (FRAGMENT) putative"
 /codon_start=1
 /protein_id="BAB23684.1"
 /db_xref="GI:12836503"
 /translation="MPTTEVPQADGGQDAGDEAEPGEKFKPKNKRKNRDYVR FTPLLVLAALVAGVLMWFLGYKAETVSVQYSGSLRNLNRHFSQDLGRRESTAFR SESAKOKMLQELVASTPLGYVYANSSVYSPGEGPLTCFFWFILDIPEYQRLTSPV VRELLVDELLNSSTLASIKYTEYDEPGLVLEASVNDIVLNSLTGCIYRYVNPQ QVLPKPDQQTSCLMHLOQPEDLMIKRLEWRVDCRDVAMVDQCNVLTLEGRLT VYGSQRPQEVLAAGSVMAVYVKKGMHSYIDPFLSVKSVAFQDCQVNLTEGRLD TQGLRTPEYSPYSPHSCSWHLTPVSLDYLALFDAYALRQKYNRLCTQGGWMI QNRRLCGFTLPVAPYADSVQVTINFTSOISLTGPGVQVYVSLYNQSDPCPGEF LCSVNGLCVPACDGIKDCPNGLDERNCVRAFQCEQDSTCISLPRVCDRQPDCLNGS DEEQCEQVPCGPTTFQCEDRSCVKNPNPECDGQSDCRDSDGDEQHCDCGLGSSRV GGTVSSEGEWPQASLQIRGHICGALIRWVITAACHFOEDSMASPKMLVFLGK MRQNSRPFGEYSFKVSLRPHYHEEDSHDVALQLDHPVYSATVPVCLPARSH FFEPCQHWITWGAGQREGGVPVNTLQKVDVQLPQDLCSEAYRYOVSPMLCAGYRK GKDKACQDGSGLVCPREPSGRWFLAGLVSNGLGCGRPNFEGVYTRVTRVINWIQVLT"

CDS

3 ArgAspSerHis.GlyAsnAlaSerProAla.....ArgThrProSerA 17
 746 AGGTCCTCCCATTTGAAGGGGCTGACCAAGCAGACCAACAGCTGCTGGT 795
 17 laGlyAla.....SerProAlaGlnAlaSerProAlaGlyThrPro 30
 796 CATCTGCAAGGCGCCGAAGACCTCATGATCAAGTGGCGGTGGAGTGAC 845
 31 ProGlyArgAlaSerProAlaGlnAlaSerProAlaGlnAlaSer..... 45
 846 CCGGGTCGATTGCAG...AGACAGGTTGGCGATGTACGACGCGAGCTGGC 892
 46ProAlaGlyT 49
 893 CCCTGGAGAAAGAGACTTATCACTCGGTCTATGGGTGCAGCCCGCAGGAA 942
 49 hrProProGlyArgAlaSer..... 55
 943 CCTGTGATGAGGTGCTGGCATCGGGCTCCGTATGCGCGGTGGTGGAA 992
 56ProAlaGlnAlaSerProAla 62
 993 AAAGGGCATGCATAGCTACTATGACCCCTTTCTCTCTCAGTGAAGTCTG 1042
 62 aGlyThrProProGlyArgAlaSerProGlyArgAlaSerProAlaGlnA 79
 1043 TGGCCTTCCAGGACTGCGCAGGTGAACCTGACACTGGAGGCGCGCTGGAC 1092
 79 laSer.....ProAlaArgAlaSerProAlaLeuAla..... 89
 1093 ACACAGGGCTTCCTCCGTACACCTACTACCCAGTTACTACTCTCCAG 1142
 90 ...SerLeuSerArgSerSerGlyArgSerSerSerAlaArgSerAla 105
 1143 TACCACACTGCTCCTGGCATCTCAGGTACCTCTCTGGACTACGGCTGG 1192
 105 aSerValThrThrSerProThrArgValThrLeuValArgAlaThrPro 122
 1193 CGCT.....CTGGTTCGATGCTACGCA 1215
 122 aGlyAlaValProIleArgSerSerProAlaArgSerAlaProAlaThr 138
 1216 CTGAGGAGGAGAGTACACCCACT...GTGTACTACGGCCAGCTGGAT 1262
 139 ArgAlaThrArgGluSerProGlyThrSerLeuProlys..... 151
 1263 GATCCAGAACAGAGGCTGTGTGCTCCGTACCTACCTCGACCATATGCTG 1312
 152PhethrTrpArgGluGlyGlnLysGlnLeuP 162
 1313 AGAGGATCCCATGGTGGCGCTCAGATGTGTCACCATCAACTTCACCTCC 1362
 162 roLeuile.....GlyCysValLeuLeuLeuAlaLeuVal 174
 1363 CAGATCTCCCTCACAGGCCCGGGTGT.....GCAAGTGT 1397
 175 ValSerLeuIleLeuLeuPheGlnPheTrpGlnGlyHisThrGlyIleAr 191
 1398 CTACAGCTTGT..... 1409
 191 gTyrLysGlu.GlnArgGluSerCysProLysHisAlaVal..... 204
 1410CAACCAATCAGACCCCTCGCTGGTGGTTCCTCTGCTGTG 1452

polyA_signal

/note="putative"

polyA_site

3030

/note="putative"

BASE COUNT

609 a 903 c 883 g 635 t

ORIGIN

```

205 .....ArgCysAspGlyValValAspCysLysLeuLy 215
1453 AATGACTGTGTCTCCCTGCGGTGACGGGATCAAGGACTCCGCCAATGG 1502
215 sSerAspGluLeuGlyCysValArgPheAspTrpAspLysSerLeuLeuL 232
1503 CCTGGATGGAGAAACGTGTGTC.....TGCAGAGCCATGTTC 1540
232 ysIleTyrSerGlySerSerHisGlnTrpLeuPro...IleCysSerSer 247
1541 AGTCCAGAGGACGACGATGCACTTCATCTGCCTAGAGTCTGTGACCGG 1590
248 .....AsnTrpAsnAspSerTyrSerGluLysThrCysGL 259
1591 CAGCCCGACTGTCATGAGTGGCAGTGAC.....GAAGAACAGTGCCA 1631
259 nGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAlaHisArgA 276
1632 AGAA...GGAGTCCCTGTGGGACATTCATCTCCAGTGTGAGGACCGG. 1677
276 spPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrIleGlnGlu 292
1678 .....AGTGTGTGAAGAAG 1692
293 SerLeuHisArgSerGluCysProSerGlnArgTyrIle..... 305
1693 CCC.....AACCCAGAGTGTGACGGCCAGTCAGATTGCCAGAGCGGCTC 1736
306 .SerLeuGlnCysSerHisCysGlyLeuArgAlaMetThrGlyArgIleV 322
1737 AGATGAGCAACACTGACTGCTGGCCCTCCAGGGCCTCTCCAGCCGTATTG 1786
322 alGlyGlyAlaLeuAlaSerAspSerLysTrpProTrpGlnValSerLeu 338
1787 TGGCGGGACCGTGTCTCCGAGGGTGAGTGGCCATGSCAGGCCAGCCTC 1836
339 HisPheGlyThrThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTr 355
1837 CAGATTGGGGTGCACACATCTGTGGGGGGCTCTCATCGCTGACCGCTG 1886
355 pValLeuThrAlaAlaHisCysPhePheValThrArgGluLysValLeuG 372
1887 GGTCTATACGGCCCGCCCTCTCCAGGAGGACAGCATGGCCCTCCCGCA 1936
372 luGlyTrpLysValTyrAlaGlyThrSerAsnLeuHis...GlnLeuPro 387
1937 AGCTGTGACCGCTGTCTCGGGAAGATGCGGCAGAACTCGCGCTGGCCA 1986
388 GluAlaAlaSer.....IleAlaGluIleIleIleAsnSerAsnTyrTh 402
1987 GGCAGGTGTCTTCAAGGTGAGCGGTGTCTGTCTGCACCGGTACACGA 2036
402 rAspGluGluAspAspTyrAspIleAlaLeuMetArgLeuSerLysProL 419
2037 GGAGACAGCATCAGTACGACGTGGCCCTGCTGCAGCTGCACCAACCCCG 2086
419 eurThrLeuSerAlaHisIleHisProAlaCysLeuPromethHisGlyGln 435
2087 TGGTGTACTCGGCACATGTGGGCCCGCTGTCCTGCTGCCGCTGCCCTCC 2136
436 ThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrAr 452
2137 TTCCTTGGCAGCGCCAGCAGCTGTGGATCACAGGCTGGGAGGCCAGCG 2186
452 gGluThrAspAspLysThrSerProPheLeuArgGluValGlnValAsnL 469
2187 AGAG...GGTGGTCCGGTGAGCAACACCCCTGCAGAGGTTGACGTACAG 2233
469 euIleAspPheLysLysCysAsnAspTyrLeuValTyrAspSerTyrLeu 485
2234 TGGTCCCTCAGGACCTCTGCAGTGAG.....GCCTACCGCTACCAAGTG 2277

```

```

486 ThrProArgMetMetCysAlaGlyAspLeuArgGlyGlyArgAspSerCy 502
2278 TCCCCAGCATGCTCTGTGTGCTACCGCAAGGCGAAGAAGATGCGTG 2327
502 sGlnGlyAspSerGlyGlyProLeuValCys...GluGlnAsnAsnArgT 518
2328 CCAGGGTGACTCTGGAGGCCACCTGGTTTGAGGGAGCCCGAGTGGCCGCT 2377
518 rPtyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsn 534
2378 GGTTCCTGGCAGGTTGTGTAGCTGGGCGCTGTGTGGCCGACCCAAT 2427
535 LysProGlyValTyrThrLysValThrGluValLeuProTrpIle 549
2428 TTCTTTGGCGTCTACACCCGTTGCACAGTGTGTATCAACTGGATC 2472
seq_name: gb_est2:BE280394

seq_documentation_block:
LOCUS BE280394 751 bp mRNA linear EST 13-JUL-2000
DEFINITION 601158674F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3505123 5',
mRNA sequence.
ACCESSION BE280394
VERSION BE280394.1 GI:9155397
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LLCMI81 row: d column: 20
High quality sequence start: 6
High quality sequence stop: 513.

FEATURES
Location/Qualifiers
source 1..751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3505123"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

alignment_scores:

Quality: 583.00 Length: 281

Ratio: 3.168 Gaps: 16

Percent Similarity: 65.480 Percent Identity: 58.719

alignment_block:

US-09-879-792-12 x BE280394 ..

Align seg 1/1 to: BE280394 from: 1 to: 751

```
1 MetGluArgAspSerHisGlyAsnAlaSerProAlaArgThrProSer.A 17
|||||
31 ATGCAGAGGGACAGCCAGCGGAATGCATCTCCAGCAAGAACACCTTCAGA 80
|||||
17 laGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArg 33
:|||||
81 CTGAGAGCTTCCAGCCAGCATCTCCAGCTGGGACACCTCCAGGCGG 130
|||||
34 AlaSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPr 50
|||||
131 GCATCTCCAGCCAG..... 145
50 oProGlyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProProG 67
145 ..... 145
67 lyArgAlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaArg 83
145 ..... 145
84 AlaSerProAlaLeuAlaSerLeuSerArgSerSerSer.GlyArgSerS 100
|||||
146 .....GCATCACTTCCAGGTCCTCATCTCCTGGCAGGTCAT 180
|||||
100 erSerAlaArgSerAlaSerValThrThrSerProThrArgValTyrLeu 116
|||||
181 CATCCGCCAGGTACCGCTCGGTGACAACTCCGCCAACAGAGTGTACCTT 230
|||||
117 ValArgAlaThrProValGlyAlaValProIleArgSerSerProAlaAr 133
|||||
231 GTTAGACAAACACCAAGTGGGGCTGTACCCATCCGATCATCTCTGCCAG 280
|||||
133 qSerAlaProAlaThrArgAlaThrArgGluSerProGlyThrSerLeup 150
|||||
281 GTCAGCACCAACCAAGGCCAGGTGGAGAGCCAGGTACGAGCCTGA 330
|||||
150 rolysPheThr...TrpArgGluGlyGlnLysGlnLeuPro.LeuIleGly 165
|||||
331 CCAAGTTCACTGACGAGCGGAGGCCAGAGCAGCTACCGACTCATCGGA 380
|||||
166 Cys....ValLeuLeuLeuIleAlaLeu.ValVal.SerLeuIleIleLeu 180
|||||
381 GTGCAGTGCTCATCTCCATGTGCCCTGGATGGTTACGCTCATCATCTC 430
|||||
181 PheGlnPheTrpGlnGlyHisThrGlyIleArg.TyrLysGluGlnArg. 196
|||||
431 TTCAGTTCTGGCAGGCGCACACAGGGATCAAGGTCACAGAGGCAAGATG 480
|||||
197 ....GluSerCysProLysHisAla.ValArg.....CysAs 207
|||||
481 TGTGAGAGCTTGTCCCAAAGCACGCCCTGTTCGCTTGTGCACGGGGTGTAT 530
|||||
207 pGlyValValAspCysLysLeuLysSerAspGluLeuGlyCysValArgp 224
|||||
531 GGGATCTTCCAAAGACTGACGAGGTGACAGCGCTGTGCTAGCGTGGAGTA 580
|||||
224 heAspTrpAspLysSerLeu...LeuLysIle.TyrSerGlySer.SerH 239
|||||
581 TTGACTGGGACCACAGCTCTGTGCTTTAAATCTTACTCTGGTCCCTCCA 630
|||||
239 lsGlnTrpLeuProIleCysSerSerAsnTrp.AsnAspSerTyrSerGl 255
|||||
631 ATCAGTGGGATCCCATCTGTAGCAGCACCTGGGAATTGACTTACTACTAC 680
|||||
255 uLysThrCysGlnLeuGly 262
|||||
681 AGAAGACTGCCAGCGAGTGGGA 702
```

seq_name: gb_htc:AK002694

seq_documentation_block:

LOCUS AK002694 1814 bp mRNA linear HTC 19-JAN-2002

DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610030A17:hepsin, full insert sequence.

AK002694
1 GI:12832864

HTC; CAP trapper.
Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
clone:0610030A17.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

10349636

2 (sites)

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

3 (sites)

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 1814)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kouda,M., Koya,S., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Kasukawa,T., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Taya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer

[5' GAGACAGAGAGCGCGCACTCGAGTGTGTTTGTGTTTGN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Seq = 5.0. Second strand cDNA was prepared with the primer adapter of sequence¹⁵, GAGACAGAGAAGTATCGAAGCTCATTAATTAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end; 3' end;

FEATURES	source
chr1: host: SDR:	Location/Qualifiers
	l. .1814
	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="MGI:1902699"
	/db_xref="taxon:10090"
	/clone="0610030A17"
	/sex="male"
	/issue_type="kidney"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"

```

gene
/ dev_stage= audit
1. .1511
/ gene="Hpn"
CDS
<1. .1511
/ gene="Hpn"
/ note="data source:MGD, source key:MGI:1196620,
evidence:ISS
hepsin
putative"

```

putative"
/codon_start=3
/protein_id="BAB22289.1"
/db_xref="GI:12832865"
/translation="RQLGRVADAGCCCHPLGCPLLMGOTPCPCGCPETNPKNPAPS
PANRPVPPNPNNSTWESLTPVDMADKEEPGAHRGSGTCSRPOKCGKGRGTAACSR
KVAALLVTLFLTGIGLAAASWALVTLLOSQDEPLYQVQLSPGDSRLAVLFDKTEGTR
LICSRSNRVAGLGEEMGFRLAHSELVDITAGANGSTGFFCVDGGLPIAQRLL
DNISVDCPRGRFLTATQCQDGRSCRWLTAGRTAVWEGGRSGACVMNGTTSVGG
PAVGLGADCTLLSRAGDSVSAICWCSPDLTGCCATQVAVIHHGYLLFRDP
TIDENSEALVNLSSLPTEYIQPVCLPAAGVALDGVKCVTVTGWNTQFYGGQAM
VLQEARVPIINSEVNCSPDFYGNQIKPKMFCAGYPEGGIDACQDGGSPVCEDSISG
TSRWRLGIVSMSTGCALRKKGVYTKVDFREWIKAIKTHSEASGMVTPD"

polyA_signal					
1977.._l802					
/note="putative"					
polyA_site					
1814					
/note="putative"					
BASE COUNT	343 a	529 c	551 g	391 t	
ORIGIN					
alignment_scores:					
Quality	577.00				Length: 514
Ratio:	1.892				Gaps: 20
Percent Similarity:	59.339				Percent Identity: 33.658

```
alignment_block:
US-09-879-792-12 x AK002694    ..
Align seg 1/1 to: AK002694 from: 1 to: 1814
```

```

93 ArgSerSerSerGlyArgSerSerSerAlaAlaArgSerAlaSerValThrTh 109
   |||||:|||||::: ||| :||||:||||:
7  CGCAGCTCGGCAGGCCACCGCTGCTGCTGCTGCTGCCACCCITGCCTT 56
   |||||:|||||::: ||| :||||:||||:
109 rSerProThrArgValrLeuValArgAlaThrProValGlyAlaValp 126
   :|||:|||||::: ||| :||||:||||:
57 CCGGGCTGTCGGCTGCTGTGGGGACAGA..... 84
   :|||:|||||::: ||| :||||:||||:
126 rolLeArgSerSerProAla.....ArgSerAlaproAlaThrArg 139
   :|||:|||||::: ||| :||||:||||:
85 .....CACCATGCCCTGCCCGCCGGGAGACTAACCCCAACCTGCAC 129
   :|||:|||||::: ||| :||||:||||:
140 AlaThrArgGluSerProGly.....ThrSerLeuPro... 150
   :|||:|||||::: ||| :||||:||||:
130 CATCTCGGGCAACCCAGGGTTCGGCCCGCCAGCCCAACAGCTCAACCTGG 179
   :|||:|||||::: ||| :||||:||||:

```

151LyspHeThrTrpArgGluCluYcGlnLysGlnLeuP 167
180 GAATCATTAACAAGAGTCCCTGCATCGCGAAGGAGATGAGGAACCTG 229
162 roLeuIleGlyCysValLeuLeuLeu.....IleAlaLeuVal 174
230 GGCTCAGAGAGGAGTTCCACTGTGTTCAAGACCCCAACCTGGAAAGGTG 279
175 valSerLeu.IleIleLeuPheGlnPheTrpGlnGlyHisThrGlyLeuA 191
280 GCGGAGTGCAGCATGCTGCTCCACAGCCCAAGGTGGCAGCTCTCATTTGTG 329
191 rgTyLySGLuInArgInGluSerCysProLySHisAlaValArgCysAsp 207
330 GGTAACCTGCTGTTCTTCAG.....AGGCATTTGGGCGCGCTCTG 370
208 Gly.ValValAspCysLysLeuLysSerAspGlu.....LeuGlyCysV 222
371 GGCCATTGTGACCATCTACTGCAGAGTACACGAGGACGCACTGTATCCAAG 422
222 alArgPheAspTrpAspLysSerLeuLeuLysIleTySerGlySerSer 238
421 TGCAGCTCAGTCCAGGGAGCTCAGGCTTCGGGTGTTTGACACAGCCGAG 470
239 HisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTySerGl 255
471 GGAACGTGGAGGTACTGCTCTCTCAGCTCCCAATGCCAGGTGGCAGG 520
255 uLysThrCysGlnLeuGlyPheGluSerAlaHisArgThrThrGluV 272
521 GCTCGGCTGTGAGGAGATGGGCTTCTCAGGGCT.....C 555
272 alAlaHisArgAspPhe.....AlaAsnSerPheSer 282
556 TGGCGCACTCGGAGCTGGATGTGCGCACTGGGGCGCCACAGCGACATCG 605
283IleLeuArgTyrAsnSerThrIleGl 291
606 GGCTTCTTTTTCGTGGACGAGGCGGAGTGCCTCTGCTCATGAGGTGCT 655
291 nGluSerLeuHisArgSerGluCysProSerGlnArgTyrIleSerLeuG 308
656 GGATGTCATCTGTATGTGACTGCTCTAGAGCGCGATCTCTGACTGCCA 705
308 lncysSerHisCysGlyLeuArg.AlamEtThrGly.ArgIleValGlyG 324
706 CTTGCCAAGACTGTGGCCGACAGAGCTGCCGGTGGACCGCATTTGGGGG 755
324 lyAlaLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPhe 340
756 GCCAGGACACAGCTCTGGGAAGGTGGCGGTGGCAGGTTCACGCTCGGTAT 805
341 GlyThrThrHisLysGlyGlyThrLeuLeuAspAlaGlnTrpValLe 357
806 GATGGGTACCACTCTCTGGGGGGTC..CTGCTGTCTCGGACGCTGGTCT 854
357 uThrAlaAlaHisCysPheValThrArgGluLysValLeuGluGlyT 374
855 GACTGTGTCACATTGC...TTTCAGAGCGGAGACAGGTCTCTGTCTCGGT 901
374 rPlysValTyAlaGlyThr.....SerAsnLeuHisGlnLeuPro 387
902 GCGGAGTATTTGCTGGGTGTGTAGCCCGGACCTCACCCCAATGCTGT... 947
388 GluAlaAlaSerIleAlaGluIleIleIleAlaAsnSerAsnTyr..... 401
948 ..GCAACTGGGTTCAGGCTGTGATCTATCATGGGGCTACCTCCCTCT 994
402ThrAspGluAspAspTyAspIleAlaLeuMetArgL 415
995 TCGAGAGCCCTCATCTACGACGAACACAGCATGATTTGGCTTGGTCAACC 1044
415 euSerLysProLeuThrLeuSerAlaHisIleHisProAlaCysLeuPro 431


```

US-09-879-792-12 x BC008514 ..
Align seg 1/1 to: BC008514 from: 1 to: 3248

19 AlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgAlaSe 35
||||| ||||||| ||| ||| |||
852 GCTTCACTACCCCGGCTTCCCA.....ACATCCCTACCCGCGCAT 895
||||| ||||||| ||| ||| |||
35 rProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro. 51
||||| ||||||| ||| ||| |||
896 GCCCGCT.....GCCAGTGGTCTCGGGGGGACCGGACTC 933

52 ..GlyArgAlaSerProAlaGlnAla..... 59
||||| ||||||| ||| ||| |||
934 TGTGCTGAGCCTCACCTTCCAGAGCTTTGATGTCGCTCCCTGTGATGAGC 983

60 .....SerProAlaGlyThrProGlyArgAlaSerPr 71
||||| ||||||| ||| ||| |||
984 ATGCCAGTGACCTGCTCAGCGTGATGATAGCCTGAGCCCATGGAACCC 1033

71 o.....GlyArgAlaSerProAlaGln..... 78
||||| ||||||| ||| ||| |||
1034 CACGCTGTGGTGGCGGTGTGGGACCTTCTCACCCCTCTACACCTGAC 1083

79 ..AlaSerProAlaArgAlaSerPro..... 86
||||| ||||||| ||| ||| |||
1084 TTTCCTCTCTCCAGACGCTCTTCTTGTGCACGCTGATAACCAATACTG 1133

87 .....AlaLeuAlaSerLeuSerArgSerSerGlyAr 98
||||| ||||||| ||| ||| |||
1134 ACCGGCAGATCCTGGCTTTGAGGCCACTTCT..... 1166

98 gSerSerAlaArgSerAlaSerValThrThrSerProThrArgValT 115
||||| ||| ||||||| ||| ||| |||
1167 .TCCAGCTGCCCAAGATGAGCAGCTGTGGCGCT..... 1199

115 yrLeuValArgAlaThrProValGly.....AlaValProIleArgSer 129
||||| ||| ||||||| ||| ||| |||
1200 .....TTTTGAGTGACACCCCAAGGACATTTAGCAGCCCTTACTATCCAG 1244

130 SerProAlaArgSerAlaProAlaThrArgAlaThrArgGluSerProG1 146
||| ||| ||| ||| ||| ||| ||| |||
1245 GCACCTACCGGCCCAACATCACTGCACATGGAATATCAAGGTGCCCAAC 1294

146 yThrSer..... 148
||||| ||| ||| ||| ||| ||| ||| |||
1295 AACCGGAAGTGAAGTGGCGCTTCAAACTCTTCTATCTGTGGACCCCAA 1344

149 .....LeuProLysPheThrTrpArgGluGlyGlnLys 159
||||| ||| ||| ||| ||| ||| ||| |||
1345 CGTACCAGTGGGTCTCGACCAAGGACTATGTGGAGATCAACGGGAGA 1394

160 GlnLeuProLeuIleGlyCysValLeuLeuLeuIleAlaLeuValValse 176
||| ||| ||| ||| ||| ||| ||| |||
1395 AGTACTCGGTGAGAGTCCAGCTTGTGTGGTGAGCAGCAAC.....AG 1437

176 rLeuIleIleLeuPheGlnPheTrpGlnGlyHis.....ThrG 189
||| ||| ||| ||| ||| ||| ||| |||
1438 CAGCAAGATTACAGTCCACTTCCATTCTGATCACTCGTACAGGACACCG 1487

189 ly.....IleArgTyrLysGluGlnArgGluSerCysPro 200
||| ||| ||| ||| ||| ||| ||| |||
1488 GGTCTCTAGTGAAGTCTCTCTAC...GACTCCACAGCCCGTGCCCA 1534

201 .....LysHisAlaValArgCy 206
||| ||| ||| ||| ||| ||| ||| |||
1535 GGGATGTTATGTCAGAGCTGACGGTGTGTCGGAAGCACTGCGCTG 1584

206 sAspGlyValValAspCysLysLeuLysSerAspGluLeuGlyCysVala 223
||||| ||| ||| ||| ||| ||| ||| |||
1585 CGACGGGTGGGCACTGCCCGGATTATAGTGTAGCGGTACTGC.... 1630

223 rgPheAspTrpAspLysSerLeuLeuLysIleYrSerGlySerHis 239

```

```

1631 .....CGATGCAATGCCACCCAC 1648
||||| ||| ||| ||| ||| ||| ||| |||
240 Gln.....TrpLeuProI1 244
1649 CAGTTACGTGCAAAAACCAAGTCTTCAAGCCCTCTTCTGG.....GT 1692
244 eCysSerSer.....AsnTrpAsnAspSerTyrSerGluLysThrCys. 258
||||| ||| ||| ||| ||| ||| ||| |||
1693 CTGTGACAGTGTCAACGACTGTGGGACGGAAGTGACGAGGAGGCTGCA 1742
258 ..... 258
1743 GCTGCTCTGCTGGGAGTTTCAAGTGTTCATGGGAAGTGTCTCCCTCAG 1792
259 .....GlnGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAl 273
1793 AGCCAGAGTGTATGGGAGGACCACTGTGGAGATGGGTCTGACGAGGC 1842
273 aHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrAsnSerThrI 290
||||| ||| ||| ||| ||| ||| ||| |||
1843 TTCATGTGACAGCGTGAATGTCGTCTCTTGCCACCAATATACCTACCGCT 1892
290 leGlnGluSerLeuHisArgSer.....GluCysProSerGln 302
||||| ||| ||| ||| ||| ||| ||| |||
1893 GCCAAATGGCTCTGTCTGAGCAAGGGCAACCCCTGAGTGTGTGGGAAG 1942
303 ArgTyrIleSerLeuGlnCysSer.....Hi 311
1943 .....ACGGACTGTAGCGATGGTCCGATGAGAAAACCTGTGA 1980
311 sCysGlyLeuArgAlaMetThr.....GlyArgIleValGlyGlyAlaL 326
||||| ||| ||| ||| ||| ||| ||| |||
1981 CTGTGGGCTCGATCCTTTACCAAAACAGGCTCGCGTGTGGTGGCAGCA 2030
326 euAlaSerAspSerLysTrpTrpGlnValSerLeuHis....PheGly 341
||||| ||| ||| ||| ||| ||| ||| |||
2031 ATCGGACGAGGGGAGTGGCCCTGGCAGGTGAGCCTCCACGCCCTGGGC 2080
342 ThrThrIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuTh 358
||||| ||| ||| ||| ||| ||| ||| |||
2081 CAGGGCCACTGTGTGGGGCTCGCTCATCTCTCTGACTGGCTGGTCTC 2130
358 rAlaAlaHisCysPhe.....PheValThrArgGluLysV 370
||||| ||| ||| ||| ||| ||| ||| |||
2131 TGACGCTCATTCCTTCAGGATGACAAAATTTCAAGTACTCAGACTACA 2180
370 alLeuGluGlyTrpLysValTyrAlaGlyThrSerAsnLeuHisGlnLeu 386
||||| ||| ||| ||| ||| ||| ||| |||
2181 CGATG.....TGGACGGCTTCTTGGGTCTG.....CTGGACCAAGGC 2218
387 ProGluAlaAlaSer.....IleAlaGluIleIleI 397
||||| ||| ||| ||| ||| ||| ||| |||
2219 AAGCGCAGTGCCTCTGGGGTGCAGGAGCTGAAGCTCAACGATATCATC 2268
397 eAsnSerAsnTyrThrAspGluGluAspAspTyrAspIleAlaLeuMetA 414
||||| ||| ||| ||| ||| ||| ||| |||
2269 CCACCTTCCTTCAATGATTTCACCTTCGACTATGACATCGCTTGTCTG 2318
414 rgLeuSerLysProLeuThrLeuSerAlaHisIleHisProAlaCysLeu 430
||||| ||| ||| ||| ||| ||| ||| |||
2319 AGTGGAGAAGTGGTGGAGTACAGCACCCCTCGTGGCGCCCATCTGCCTG 2368
431 ProMetHisGlyGlnThrPheSerLeuAsnGluThrCysTrpIleThrG1 447
||||| ||| ||| ||| ||| ||| ||| |||
2369 CTTGATCTACCATGCTCTCCCTGTGGCAAGGCCATCTGGGTACAGG 2418
447 yPheGlyLysThrArgGluThrAspAspLysThrSerProPheLeuArg 464
||||| ||| ||| ||| ||| ||| ||| |||
2419 CTGGGGCACACAAAGAG...GGAGTACGGGAGCGCTGATCTGTGCAGA 2465
464 luValGlnValAsnLeuIleAspPheLysLysCysAsnAspTyrLeuVal 480
||| ||| ||| ||| ||| ||| ||| |||

```


Percent Similarity: 54.839 Percent Identity: 29.677

alignment_block:

US-09-879-792-12 x AK005546 ..

Align seg 1/1 to: AK005546 from: 1 to: 2246

```
193 LysGluGlnArgGluSerCysProLysHisAlaValArgCysAspGlyVa 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
595 AAGTACACCGAAGGGGACGCCAACACCAATAACGAAGCTCAATGGCGT 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 lvalasp...CysLysLeuLysSer.....AspGluLeuGlyCysV 222
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
645 GGTATCTGGATTTTCACTGAAGTCTGTGGAGCTTTCAAACTTGGCTGTA 694
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 alArgPheAspTrpAspLysSerLeuLysIleTyrSerGlySerSer 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
695 TCAGGGACATTTTCCCTAACACGGTGTGGCAGACCTTAACATGACAG 744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
239 ..... 241
745 GTGTGCCCCAGATGCTTTTGTCTGTCTGTCGTCATCTGCACGCATCAC... 792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 pLeuProIleCys.....SerSerAsnTrpAsnAsps 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
793 ....CCACATGTTGTGTCTTCATCTTTTCCCAAGCATGGCCGAAG 838
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 erTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArg 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
839 AATCTCAGACACATCTTGTCTCTTAAACCTCTGAAAGTGGATACCA 888
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
269 ThrThrGluValAlaHisArgAspPheAlaAsnSerPheSerIleLeuAr 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
889 AGCACAGCATTTACAAAGAGCCAGCCCTTTCGGGCTTCAGTCTCCAGCA 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 g.....TyrAsnSert 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
939 CTGCAGGCACAGTGTCCAGTATTCTGCCATCGTCTCTTTTACAACGACA 988
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
289 hr.....IleGlnGluSerLeu..... 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
989 CTGATTTCTTGGGAGAGAGCTGGACATCTCGATGTGAAAGGCCAAGAA 1038
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
295 .....HisArgSerGluCys..... 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1039 ACCTGTGAGAAAACGTGTACCAATAAGCCCGCTGCCAGTCTTTCACCTA 1088
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 ....ProSerGlnArg.....T 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1089 CTATCCATCGCACAGACTGTGCAATGAGAGGAACCCGAGGGGAGATGTT 1138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 yrIleSerLeuGlnCysSerHisCysGlyLeuArgAlaMetThrGly... 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1139 ACCTAAGCTTCTCTCAATGGATCTCCACAGAGAAATACCTTCATGGGAG 1188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
319 ..... 319
1189 GGAGGCATCTCTGGATACTCACTGAGGCTGTGCAAAATGGATAATGTGTG 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 .....ArgIleValGlyGlyAlaLeuAlaSerAsps 330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1239 CACAACCTAAATCAACCCAGAGTGTAGGAGGCTGCTCTGTTCACG 1288
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 erLysTrpProTrpGlnValSerLeuHisPheGlyThrThrHisIleCys 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1289 GTGAGTGGCCATCGAGGTGACTCTGCATCAGCAGCCAGGACCCCTGTGT 1338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 GlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPh 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1339 GGAGGCTCCATCATTTGGAACCAATGGATCTACAGCAGCAGCTCATGTTT 1388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
363 e.....PheValThrArgGluLysValLeuGluGlyTrpLysValTyrA 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
1389 CTCCTGGGATAGAGACACCTAAAAAGCTG.....CGTGTCTACG 1426
378 laGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSer..... 391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1427 GTGCATTTGTAATCAATCAGAAATAATGAAGGACTGCTTTCTTCAGG 1476
392 ITeAlaGluIleIleAsnSerAsnTyrThrAspGluGluAspAspTyr 408
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1477 GTTCAAGAATGATAATTCATGATCAGTATACGACAGCAGAGAAGTGGTA 1526
408 rAspIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisI 425
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1527 TGATATTGCCCTGTAAACTGGAATCAGCCATGATTTACACAGATTTTC 1576
425 leHisProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsnGlu 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1577 AGCGGCCAATATGCTGCTTCCAAAGAGATAGAAACGACGAGTGCACACA 1626
442 ThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLysTh 458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1627 GAATGCTGGGTGACTGGATGGGGTACACAGCAGCTAAGAGGTGAAGTACA 1676
458 rSerProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysC 475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1677 AAGTACT...CTCAGAAAGCAAGGTTCATTTGCTGTCAAATGAAGAAT 1723
475 ysAsnAspTyrLeuValTyrAspSerTyr...LeuThrProArgMetMet 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1724 GTCAGACAAGA.....TACAGAAGACACAAAATAACCAATAAGATGATC 1767
491 CysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGl 507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1768 TGTGCAGCTACAAAGAGGAGGAGGATACGTGCAAGGAGATTTCTGG 1817
507 yGlyProLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValT 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1818 AGGCCCTCTGCTCTGCAAAATACATACATGGGTCTGGCAGTGGTGGCATCA 1867
524 hrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThr 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1868 CAAGCTGGGTGAAGGCTGTGTCAGAGAGAGAGACCGGGGGTCTACACG 1917
541 LysValThrGluValLeuProTrpIleTyrSerLysMetGluSer 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1918 AACGTGCCCAAGTACGTGGACTGGATTTCTGAGAGAAAACCTCAACA 1962
```

seq_name: gb_est2:BE309103

seq_documentation_block:

LOCUS BE309103 688 bp mRNA linear EST 26-OCT-2000
DEFINITION 601093231F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3488047 5',
mRNA sequence.

ACCESSION BE309103 GI:9166776

VERSION BE309103.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 688)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LLAM8526 row: m column: 08

```
FEATURES
Source
High quality sequence stop: 457.
Location/Qualifiers
1. .688
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH108"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 152 a 210 c 171 g 155 t
ORIGIN
alignment_scores:
Quality: 529.00 Length: 212
Ratio: 3.187 Gaps: 19
Percent Similarity: 78.302 Percent Identity: 66.509
alignment_block:
US-09-879-792-12 x BE309103 ..
Align seg 1/1 to: BE309103 from: 1 to: 688
179 ILeuLeuPheGlnPheTrpGln...GlyHisThrGlyIleArgTyrLysG1 194
:::||||| |||||::: |||||
19 GTTCTCTCTACTTCTGGAGAGTGCCACACGCGGATCAAGTACAACGGA 68
194 uGlnArgGluSerCysProLysHisAlaValArgCysAspGlyValVala 211
| |||||
69 GCCACTGAGAGTGCCTTCCACACGCGGTTCGCTGTGATGGAGTGTGG 118
211 sPCysLysLeuLys.SerAsp.GluLeuGlyCysValArgPheAspTrpA 227
|||||
119 ACTGCAAAATGAAGCAGCATAGAGTGGGCTGTGTTCAGGTTCGACTGGG 168
227 sPlysSerLeuLysIleTyrSerGlySerSerHisGlnTrpLeuPro 243
|||||
169 ACAAGTCCCTCCTGAAAGTCTACTCTGGGCTTCTGCGAGTGGCTTCCT 218
244 ILeCysSerSer.AsnTrpAsnAspSerTyrSerGluLysThrCysGlnG 260
:::||||| :::|||||::: |||||
219 GTCTGCAGCAGCAGCTGGAGACGACACTGACTCCAGAGGACCTGCCAGC 268
260 InLeuGlyPhe.GluSerAlaHisArgThrThrGluValAlaHis.ArgA 276
|||||
269 AGCTGGCATTCGACAGCGCTTACCGAACAACACTGAGTAGCCCACTAGAG 318
276 sp.PheAlaAsnSerPheSerIleLeuArg.TyrAsnSerThr.IleGln 291
|| :::::||||| ||||| |||||
319 ACTGTACACAGCAGCTTCTTGACTCTCCGATACGACACCAACCAATCCAG 368
292 GluSerLeuHisArgSerGluCysProSerGlnArg...TyrIleSerLe 307
|||||
369 GAAAGCCTCTACAGTCCGAATATCTTCCCGGCGGTAAATGGTCTCCCAT 418
307 uGlnCysSerHisCysGlyLeuArgAla.MetThrGlyArgIleValGly 323
|||||
419 CCAGTGTTCCTGCTGTTGAGAGCTATGACCGGCGGATGCTGGGA 468
324 Gly.AlaLeu.AlaSerAspSerLysTrp.ProTrpGln.ValSerLeuH 339
|| |||||
469 GGGGCTCTGAACCTCGGAGAGCAAGTGCCTCGCTAAGTTAGCCTGC 518
339 IsPheGly.ThrThrHisIle.CysGlyGlyThrLeu.IleAsp.AlaG1 354
|||||
519 ACTTCGGCACTACCCACATTCGTGTGGCGGCACACTTCATCGATAGCCCA 568
|||||
```

```
354 nTrpValLeuThrAlaAlaHisCysPheValThrArgGluLysValL 371
| |||||
569 GTGTGTCTCCACCGGTGCCACCGTTTTTTGTGACCCCGCAACACCTC 618
371 eu 371
||
619 TT 620
seq_name: gb_est1:AL555870
seq_documentation_block:
LOCUS AL555870 876 bp mRNA linear EST 16-FEB-2001
DEFINITION AL555870 LTI_NFL006.PL2 Homo sapiens cDNA clone CSODK002YM07 5
prime, mRNA sequence.
ACCESSION AL555870
VERSION AL555870.1 GI:12898012
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. .876
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODK002YM07"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 171 a 251 c 266 g 188 t
ORIGIN
alignment_scores:
Quality: 513.50 Length: 256
Ratio: 2.951 Gaps: 6
Percent Similarity: 67.969 Percent Identity: 41.016
alignment_block:
US-09-879-792-12 x AL555870 ..
Align seg 1/1 to: AL555870 from: 1 to: 876
312 CysGlyLeuArgAlaMetThrGlyArgIleValGlyAlaLeuAlaSe 328
|||||
104 TGGCGCGCAGCGGTTCATCAGTCGCGCATCTGGTGGAGAGGACGCGGA 153
328 rAspSerLysTrpProTrpGlnValSerLeuHisPheGlyThrThrHisI 345
|| :::::||||| |||||
154 ACTCGGCGGTGGCCGTGGCAGGGGAGCCTGCGCCTGGGATTCACAG 203
345 leCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHis 361
:::||||| :::::|||||
204 TATGCGGAGTGAGCTCTGCTCAGCCACCGCTGGGCACTCAGCGGCGGCAC 253
```

```
362 CysPheValThrArgGlu...LysValLeuGluGlyTrpLysValTy 377
||||| : : : : : : : : : : : : : : : : : :
254 TGTCTTGAACCTATAGTACCTTAGTATCCCTCCGGGTGGATGGTCCA 303
||||| : : : : : : : : : : : : : : : : : :
377 rAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSer..... 391
||| : : : : : : : : : : : : : : : : : :
304 GTTTGGC.....CAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGG 347
392 .....IleAlaGluIleIleAsnSerAsnTyr 401
: : : : : : : : : : : : : : : : : :
348 CCTACTACACCCGTTACTTCGTATCGAATATCTATCTGAGCCCTCGCTAC 397
402 ThrAspGluGluAspTyrAspIleAlaLeuMetArgLeuSerLysPr 418
: : : : : : : : : : : : : : : : : :
398 CTGGGG...AATTCACCTATGACATTGCCTTGGTGAAGTGTCTGCACC 444
418 oLeuThrLeuSerAlaHisIleHisProAlaCysLeuPromethHisGlyG 435
| : : : : : : : : : : : : : : : : : :
445 TGTCACTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACAT 494
435 lnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThr 451
||||| : : : : : : : : : : : : : : : : : :
495 TTGAGTTTGAGACCGGACAGACTGCTGGGTGACTGCTGGGGGTACATC 544
452 ArgGluThrAspAspLysThrSerProPhe...LeuArgGluValGlnVa 467
: : : : : : : : : : : : : : : : : :
545 AAAGAGGATGAGGCACCTGCCATCTCCCCACACCCCTCCAGGAAGTTCAGGT 594
467 lAsnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrAspSerT 484
| : : : : : : : : : : : : : : : : : :
595 CGCATCATAAACACTCTATGTGCAACACCCTCTTCTCAAGTACAGTT 644
484 yr.....LeuThrProArgMetMetCysAlaGlyAspLeuArgGly 497
: : : : : : : : : : : : : : : : : :
645 TCGCAAGGACATCTTTGGAGACATGTTTGTGTGGCAATGCCCAAGGC 694
498 GlyArgAspSerCysGlnGlyAspSerGlyGlyProLeuValCysGluGl 514
||||| : : : : : : : : : : : : : : : : : :
695 GGGAAAGGATGCCGTCTCGGTGACTCAGGTGGACCCCTTGGCCCTGTAACA 744
514 nAsnAsnArgTrpTyrLeuAlaGlyValThrSerTrpGlyThrGlyCysG 531
: : : : : : : : : : : : : : : : : :
745 GAATGGACTGTGTATACAGATTGGAGTCGTGAGCTGGGGAGTGGGCTGTG 794
531 lyGlnArgAsnLysProGlyValTyrThrLysValThrGluValLeuPro 547
| : : : : : : : : : : : : : : : : : :
795 GTCGGCCCAATCGGCCGGTGTCTACACCAATATACGCCACCACTTTGAG 844
548 TrpIleTyrSerLysMet 553
||||| : : : : : :
845 TGGATCCAGGAAGCTGATG 862
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:36:49 ; Search time 63.14 Seconds
(without alignments)
988.652 Million cell updates/sec

Title: US-09-879-792-12
Perfect score: 2999
Sequence: 1 MERDSHGNSPARTPSAGAS.....TEVLPWIYKMESEVFRKS 562

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2991	99.7	562	22	AAE01943 Human transmembran
2	2963	98.8	688	22	AAE01944 Human transmembran
3	2287	76.3	421	22	AAE05042 Human SER6 protein
4	2207	73.6	491	22	AAE03442 Human protein sequ
5	1311	43.7	296	21	AAE02108 Human serine prote
6	1311	43.7	372	21	AAE02092 HRPcA6/7 polypepti
7	741.5	24.7	492	21	AAE02050 Human prostate CDN
8	741.5	24.7	492	22	AAU06960 P1000C amino acid
9	741.5	24.7	492	22	AAE01315 Human tumour suppr
10	740.5	24.7	492	21	AAE07726 Human P0P1F12-GTC2
11	740.5	24.7	492	21	AAE04406

12	736.5	24.6	492	21	AAE036901 Human TMPRSS2 prot
13	736.5	24.6	492	21	AAE057280 Ovr115 homolog pro
14	736.5	24.6	492	22	AAU069939 Human transmembran
15	736.5	24.6	492	22	AAE06943 Human serine prote
16	736.5	24.6	492	22	AAE06943 Human serine prote
17	731.5	24.4	453	21	AAE04250 Human transmembran
18	729.5	24.3	452	20	AAE041694 Human PRO382 prote
19	729.5	24.3	453	22	AAU02055 Human PRO382 prote
20	729.5	24.3	453	22	AAE06935 Human membrane-typ
21	727.5	24.3	393	22	AAE06962 Human prostate CDN
22	727.5	24.3	393	22	AAE06962 Human prostate CDN
23	726	24.2	454	21	AAE01317 P1000C partial ami
24	726	24.2	454	21	AAE01317 Tumour associated
25	726	24.2	454	22	AAE06911 Human TAGD-12, pro
26	697.5	23.3	423	22	AAE06944 Human transmembran
27	696.5	23.2	435	20	AAE06437 Human seripancrin
28	694	23.1	432	21	AAE06437 Human protease HUP
29	694	23.1	432	21	AAE06437 Human PRO1570 (UNQ
30	694	23.1	432	22	AAU029188 Human PRO polypept
31	694	23.1	432	22	AAE06944 Human PRO1570. Ho
32	692.5	23.1	492	22	AAE06931 Human seripancrin
33	692.5	23.1	492	22	AAE06931 Human seripancrin
34	684.5	22.8	521	22	AAU16975 Human novel secret
35	664	22.1	445	21	AAU16975 Mouse serine prote
36	646	21.5	480	21	AAE08950 Human secreted pro
37	644.5	21.5	457	21	AAE08950 Human serine prote
38	632.5	21.1	414	21	AAE08912 Human secreted pro
39	630.5	21.0	327	21	AAE072093 Human serine prote
40	622	20.7	317	22	AAE025633 Human protein sequ
41	613.5	20.7	416	20	AAE043325 Mouse hepsin prote
42	619.5	20.7	416	20	AAE06812 A mouse serine pro
43	611.5	20.4	417	22	AAE06812 Human hepsin prote
44	607.5	20.3	311	21	AAE01697 Mouse serine prote
45	606	20.2	273	21	AAE01696 Mouse serine prote

ALIGNMENTS

RESULT 1
ID AAE01943 standard; Protein; 562 AA.
AC AAE01943;
DT 31-JUL-2001 (first entry)
DE Human transmembrane serine protease (Endothelias 2-S) protein.
KW Human; endothelias 2-S; protease domain; cytosolic; vulnary; wound;
KW nontropic; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
KW liver cirrhosis; osteoradionecrosis; systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;
KW transmembrane serine protease.
OS Homo sapiens.

Key	Location/Qualifiers
Domain	321..562
Label	Protease_domain
WO200136604-A2	
25-MAY-2001	
17-NOV-2000; 2000WO-US31803	
18-NOV-1999; 99US-0156391	
22-SEP-2000; 2000US-0234840	

(CORV-) CORVAS INT INC.
Madison EL, Ong EO;
WPI: 2001-336001/35.
N-PSDB; AAD05796.
New nucleic acid encoding a protein comprising endothelase activity useful in the prevention and treatment of e.g. vascular malformations, cardiovascular disorders, and chronic inflammatory disease -
Claim 4; Page 138-139; 152pp; English.
The present sequence is human short form transmembrane serine protease (Endothelase 2-S) protein. Endothelase 2 is a type-II membrane-type serine protease which has a transmembrane domain at the N-terminus, followed by a single low density lipo-protein-A receptor domain and a single scavenger-receptor cysteine-rich domain. The C-terminus of the endothelase 2 contains the trypsin-like serine protease domain characterised by the catalytic triad residues in 3 highly conserved regions of the catalytic domain. In addition 3 repetitive sequence are found just before the transmembrane domain and represent a sequence motif for N-myristoylation modification.
The invention relates to an endothelase protein, endothelase protease domain and their corresponding nucleic acid molecules. An endothelase protein or protease domain of it is useful for the treatment and diagnosis of disorders associated with aberrant angiogenesis or undesired neovascularisation. The undesired angiogenesis is associated with disorders selected from solid neoplasm, vascular malformations and cardiovascular disorders such as angiofibroma, angiolipoma, atherosclerosis, restenosis/reperfusion injury, arteriovenous malformations, haemangiomatosis and vascular adhesions, dyschondroplasia with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau syndrome, chronic inflammatory diseases such as diabetes mellitus, haemophilic joints, inflammatory bowel disease, nonhealing fractures, periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers, granulomatous-burns, hypertrophic scars, liver cirrhosis, osteodysplasia, postoperative adhesion, pyogenic granuloma and systemic sclerosis and aberrant wound repairs, circulatory disorders Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal, dyomeitoly, sclerodactyly and teangiectasis, dermatological disorders such as systemic vasculitis, scleroderma, pyoderma gangrenosum, vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome and Osler-Weber-Rendu syndrome and ocular disorders such as blindness caused by ocular neovascular disease, corneal graft neovascularisation, macular degeneration, retinopathy of prematurity, retrolental fibroplasia and corneal neovascularisation. The nucleic acids of the invention are also used in gene therapy. The invention also provides method for screening compounds that modulate angiogenesis.

Query Match 99.7%; Score 2991; DB 22; Length 562;
Best Local Similarity 99.8%; Pred. No. 9.2e-189;
Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MERDSHGNSPARTSAGASPAQASPAQASPAQASPAQASPAQASPAQASPAQAS 60
Db 1 merdshgnsparpsagapagcpggraspapagapagcpggraspapag 60
QY 61 PAGTPPGRASPAQASPAQASPAQASPAQASPAQASPAQASPAQASPAQAS 120
Db 61 pagtppgraspapagapagapagapagapagapagapagapagapag 120
QY 121 PVGAVPIRSSPARSPATRATRESPTSLPKFTWREGQKQLPGCVLLIALVSLIIL 180
Db 121 pvgavpirtsparsapatratresptslpkftwregkqlplgcvllialvslil 180
QY 181 FQFWOCHTIRYKEQRESCPKHVRCDGVVDCCKLSDELGCVRFDWCKLSLLIYSSSHQ 240

Db 181 fqwghgtgirykeqrescpkhavrcdgvvdccklsdelgcvrfdwcklslliyssshq 240
QY 241 WLPICSSNNDVSEKTCOOLGFESAHRTTEVAHRDFANSFSLRYNSTIQESLHRECP 300
Db 241 wlpicsnndvsektcqlglfesahrtevhahrdfansfslrynstiqeslhrsecp 300
QY 301 SORYISLQCSHGCLRAMTGRIVGGALASDKWPQVSLHFGTHICGGTLIDAQWVLTAA 360
Db 301 sqryislqschgclramtgrivggalasdkwpqvsllhfgthicggtlidaqwvltaa 360
QY 361 HCFYFTRKXVLEGWKVVYAGTSLNHLQHPAASIAEIIINSNYTDEDDYDIALMRLSKPLT 420
Db 361 hcfyfvtrkxvlekwkvayagtsnlhlqipeaasiaeiiinsnytdeeddydialmriskplt 420
QY 421 LSAHLHPACLPWHGQTFSLNETCWTGFKTRTDKTSFPLREVQVNLIDFKKNDYLV 480
Db 421 lsaahlpacpwhgqtfslnetcwitgfktrtdktsfplrevqvnidfkkcndylv 480
QY 481 YDSYLTPRMCMAGDURGDRSCQSGGGLVCEQNNRWYLAGVTSWGTGCGGRNKPQVYT 540
Db 481 ydsyltprmmcagdlrgrdscqsggglvceqnnrwyagvtswgtcggrnkpqvyt 540
QY 541 KYTEVLPWIYSKMESEVRERKS 562
Db 541 kvtevpwiyskmesevrerks 562
RESULT 2
AAE01944 ID AAE01944 standard; Protein; 688 AA.
XX AC AAE01944;
XX DT 31-JUL-2001 (first entry)
XX DE Human transmembrane serine protease (Endothelase 2-L) protein.
XX KW Human; endothelase 2-L; protease domain; cytostatic; vulnery; wound; neotrophic; periodontitis; dermatological disorder; gene therapy; scar; anglogenesis; cardiovascular disorder; psoriasis; neovascular disease; chronic inflammatory disease; ocular disorder; circulatory disorder; crest syndrome; atherosclerosis; haemangiomatosis; diabetes mellitus; liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal; inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy; systemic vasculitis; scleroderma; neoplasm; ulcer; burn; transmembrane serine protease.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 321..688
XX FT /label= protease_domain
XX PN WO200136604-A2.
XX PD 25-MAY-2001.
XX PF 17-NOV-2000; 2000WO-US31803.
XX PR 18-NOV-1999; 99US-0166391.
XX PR 22-SEP-2000; 2000US-0234840.
XX XX
XX PA (CORV-) CORVAS INT INC.
XX XX
XX PI Madison EL, Ong EO;
XX XX
XX DR WPI: 2001-336001/35.
XX DR N-PSDB; AAD05797.
XX PT New nucleic acid encoding a protein comprising endothelase activity useful in the prevention and treatment of e.g. vascular malformations, cardiovascular disorders, and chronic inflammatory disease -
PT

CC represents the amino acid sequence of human SER6 protein.

XX Sequence 421 AA;

Query Match 76.3%; Score 2287; DB 22; Length 421;
Best Local Similarity 99.8%; Pred. No. 1.3e-142;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 143 ESPGTSLPFTWREGQKPLIGCVLLIALVSLILFQFWOGHTGIRYKQRESCPKH 202
Db |||||
QY 203 AVRCDDGVCKLSDDELGCVRFDWDSLKIYSGSHQWLPICSSNWNDSYEKTCQQLG 262
Db |||||
QY 263 FESAHTTEVAHRDFANSFSLRYNSTIOESLHRSRCPQRYISLQCSHCGLRAMTGRIV 322
Db |||||
QY 323 GGALASDSKWPQVSLHFGTHICGGLTIDAQWVLTAAHCFVTRKVLGKWKVYAGTSN 382
Db |||||
QY 383 LHQLPAASTAEIIINSNTYDEDDYDIALMRLSKPLTISAHHPACLPNHGQTFSLNET 442
Db |||||
QY 443 CWTGFGKTRTDKTSPLREVOVNLIDFKKNDYLVYDSYLTPRMCGAGDLRGGRDSC 502
Db |||||
QY 503 QGDSGGPLVCEQNRYLAGVTSWGTGCGQRNKPQVYTKVTEVLPWIYKMESEVFRKS 562
Db |||||

RESULT 4

ID AAB93442 standard; Protein; 491 AA.

AC AAB93442;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12682.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 491 AA;

Query Match

Best Local Similarity

Matches 426;

Conservative

Mismatches

Indels

Gaps

Length

491;

Score 2207;

DB 22;

Pred. No. 2.9e-137;

0;

7;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

10;

XX 28-MAR-2001 (first entry)
XX Human serine protease #3 fragment from HATEE38 cDNA clone.
DE Human; serine protease; osteopathic; immunosuppressive; antiallergic;
KW antiinflammatory; cytostatic; cardiac; neuroprotective; nootropic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening.
XX Homo sapiens.
XX WO200068247-A2.
XX 16-NOV-2000.
PF 05-MAY-2000; 2000WO-US12207.
XX 07-MAY-1999; 99US-0133239.
PR 20-MAY-1999; 99US-0135163.
PR 03-AUG-1999; 99US-0147005.
PR 09-SEP-1999; 99US-0152935.
PR 01-NOV-1999; 99US-0162979.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Shi Y, Young PE, Ni J;
PI WPI; 2000-679799/66.
DR N-PSDB; AAD02322.
XX New nucleic acid molecules encoding human serine protease polypeptides,
PT useful for diagnosis, prevention and/or treatment of disorders e.g.
PT osteoporosis, lupus erythematosus and Alzheimer's -
PS Disclosure; Page 286-287; 289pp; English.
XX The present sequence is human serine protease #3 fragment from clone
CC HATEE38 (ATCC Deposit No: PTA27).
CC The invention relates to human serine proteases and their cDNA clones.
CC It is used in methods for the diagnosis, prevention and treatment of
CC various disorders related to serine protease such as bone formation
CC disorders (osteoporosis), connective tissue disorders (arthritis),
CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
CC reproductive system disorders (testicular cancer), digestion and food
CC absorption disorders (Crohn's disease), neurodegenerative diseases
CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
CC proliferative and cancerous conditions (acute myelogenous leukaemia),
CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
CC ocular disorders (glaucoma), and infectious diseases caused by bacteria,
CC viruses, fungi or parasites. It is also useful for screening therapeutic
CC compounds. Serine proteases are used as immunological probes or
CC polymorphic markers for the identification of chromosomes, cells and
CC tissues in biological samples, identification of male contraceptive
CC agents, delivery of compositions to targetted cells expressing a
CC receptor for serine protease, hybridisation probes and molecular weight
CC markers. Serine protease nucleic acids are also useful in gene therapy.
CC Note: The present sequence shown in page 286-287 of sequence listing has
CC been assigned SEQ ID NO: 28. But the sequence, peptide fragment #25
CC related to human serine protease (AA72128) shown in page 22 is also
CC referred as SEQ ID NO: 28.
XX Sequence 296 AA;

Query Match 43.7%; Score 1311; DB 21; Length 296;

Best Local Similarity 98.3%; Pred. No. 1.4e-78;
Matches 238; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 317 MTRGVGGALASDSKWPQVSLHFGTTHICGGTLDIAQWVLTAAHCFVTRKVLGSKV 376
DB 1 mtrgrivggalsdskwpqvgvshfgtthicggclidaqvwltaaahcfvtrkvlegkw 60
QY 377 YAGTSLNHLQPEAAASIAETIIINSNTDEDDYDIALMRLSKPLTLSAHIHPACLPHGQT 436
DB 61 yagtslnhlpeaasiaeiiinsntdeeddydialmriskpltsahihpaelphbgqt 120
QY 437 FSLNETCWTGTGCKTRETDDKTSPLREYQVNLIDPKKNDYLVDSYLTPRMCAQDLR 496
DB 121 fslnetctwtgtgktrtdktsplrevqvnldfkkcndylvdsytlprmmcagdlr 180
QY 497 GGRDSCQSGSGPLVCEQNNRWYLAGVTSWGTGCGORNKPGVYTKVTEVLPWTYSKWSE 556
DB 181 ggrdscqsgsgplvceqnnrwylagvtswgtgcggrnkpgvytkvtevlpwyskmenr 240
QY 557 VR 558
DB 241 aq 242
RESULT 6
AA72092
ID AA72092 standard; Protein; 372 AA.
AC AA72092;
XX 28-MAR-2001 (first entry)
DE Human serine protease #3 encoded by clone HATEE38.
XX Human; serine protease; osteopathic; immunosuppressive; antiallergic;
KW antiinflammatory; cytostatic; cardiac; neuroprotective; nootropic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening.
XX Homo sapiens.
OS WO200068247-A2.
PN 16-NOV-2000.
PD 16-NOV-2000.
XX 05-MAY-2000; 2000WO-US12207.
XX 07-MAY-1999; 99US-0133239.
PR 20-MAY-1999; 99US-0135163.
PR 03-AUG-1999; 99US-0147005.
PR 09-SEP-1999; 99US-0152935.
PR 01-NOV-1999; 99US-0162979.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Shi Y, Young PE, Ni J;
PI WPI; 2000-679799/66.
DR N-PSDB; AAD02322.
XX New nucleic acid molecules encoding human serine protease polypeptides,
PT useful for diagnosis, prevention and/or treatment of disorders e.g.
PT osteoporosis, lupus erythematosus and Alzheimer's -
PS Disclosure; Page 286-287; 289pp; English.
XX The present sequence is human serine protease #3 fragment from clone
CC HATEE38 (ATCC Deposit No: PTA27).
CC The invention relates to human serine proteases and their cDNA clones.
CC It is used in methods for the diagnosis, prevention and treatment of
CC various disorders related to serine protease such as bone formation
CC disorders (osteoporosis), connective tissue disorders (arthritis),
CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
CC reproductive system disorders (testicular cancer), digestion and food
CC absorption disorders (Crohn's disease), neurodegenerative diseases
CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
CC proliferative and cancerous conditions (acute myelogenous leukaemia),
CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
CC ocular disorders (glaucoma), and infectious diseases caused by bacteria,
CC viruses, fungi or parasites. It is also useful for screening therapeutic
CC compounds. Serine proteases are used as immunological probes or
CC polymorphic markers for the identification of chromosomes, cells and
CC tissues in biological samples, identification of male contraceptive
CC agents, delivery of compositions to targetted cells expressing a
CC receptor for serine protease, hybridisation probes and molecular weight
CC markers. Serine protease nucleic acids are also useful in gene therapy.
CC Note: The present sequence shown in page 286-287 of sequence listing has
CC been assigned SEQ ID NO: 28. But the sequence, peptide fragment #25
CC related to human serine protease (AA72128) shown in page 22 is also
CC referred as SEQ ID NO: 28.
XX Sequence 296 AA;

QY 419 LTLSAHIHPACLPMHGQTFSLNETCWITGFGKTRTDDKTSPPFLREVQVNLIDFKKNDY 478
Db 355 ltfndlvkvpclpnpqmmllqpeqlcwisgwgatee-kgktsevinnaakvllietqrcnsr 413
QY 479 LVYDSYLTPRMCAAGDLRGGRDSCQDGGPLVCEQNNRWYLAGVTSWGTGCGQRNPGV 538
Db 414 yvydnlitpamicagflqgnvdsqgdspplvtsnnnllwllgdtswgscakayrpgv 473
QY 539 YTKVTEVLPWIYSKMS 555
Db 474 ygnvmvftdwlyrqmka 490
RESULT 14
AAU69939
ID AAU69939 standard; Protein; 492 AA.
XX
AC AAU69939;
XX
DT 30-JAN-2002 (first entry)
DE Human transmembrane serine protease 2.
XX
KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 13-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skelky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
DR N-PSDB; AAS64164.
XX
XX New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer -
XX Claim 2; Page 556-557; 579pp; English.
XX
XX The invention relates to isolated prostate-specific
XX polynucleotides, polypeptides, fusion proteins of the polypeptides,
XX antibodies raised against the polypeptides (or antigenic epitopes
XX derived from them) and antigen-presenting cells expressing the
XX polypeptides. The antibodies are useful for detecting the presence of
XX cancer, especially prostate cancer. The polypeptides, polynucleotides and
XX the antigen-presenting cells are useful for stimulating and/or expanding
XX T cells specific for a tumour protein, and for inhibiting the development
XX of cancer especially prostate cancer. Compositions comprising the
XX polynucleotide and/or polypeptide are useful for stimulating an immune
XX response, and for treating cancer. The oligonucleotide is useful for
XX detecting cancer. The present sequence is a prostate specific
XX polypeptide of the invention.
SQ Sequence 492 AA;

Query Match 24.6%; Score 736.5; DB 22; Length 492;
Best Local Similarity 32.0%; Pred. No. 1.5e-40;
Matches 178; Conservative 86; Mismatches 194; Indels 99; Gaps 18;
QY 27 AGTTPGPA-----SPAQASPAQASPAQPPGPRASPAQASPAQASPAQPPGPRASPAQ 78
Db 5 sgspaiqgyyenhyqpenypaqptvvpt-vyevhpaqyypspvp-----qyaprv 56
QY 79 ASPARASPALASLSRSSSSRSASASVTTSPTRVILVRATPVGAVPIRSSPARSAPAT 138
Db 57 ltga-snpvctqkpspg-----tvtskktkalcitltlgtf----- 94
QY 139 RATRESPGTSLPKFTWREGQKQLPLIGCVLLLLIALVSLILFQWQ-----GHTGIRY 192
Db 95 -----lvga-----alaagll-----wkfmgsksnsגיע 120
QY 193 KEQRESCPKHAVRCDDVVDCVKLSDELGCVRFDWDKSLKLIYSSSSHOWLPICSSNNWDS 252
Db 121 -dssgtclnpsnwcgvshcpggedenrcvrllygpnfilqmyssqrkshwpcvqddwnen 179
QY 253 YSEKTCQQLGFESAHRTVEAHRDFANSFSLRYNST-----IQESLHRSE-CPQSQRVIS 306
Db 180 ygraacrdmgyknnfyssq-givddsgstsfmkintsagnvdiykklyhsdacskaavv 238
QY 307 LOCCHCGLR---AMTGRIVGGALASDSKWPQVSLHFGTTHICGTLIDAQWLTAACHF 363
Db 239 lrclacgvnlinsrqsrivggesalpgawpqvshlvqnvhvcgsiitpewivtaahcv 298
QY 364 FVTRKVLG---WKVYAGTSLNLHQLPEAA--STAEIINSNYTDEEDDYDIALMRSLKP 418
Db 299 ---ekplnnpwhwtatagilrqsfmfygagvqkvqkshpnydsktknndialmklqkp 354
QY 419 LTLSAHIHPACLPMHGQTFSLNETCWITGFGKTRTDDKTSPPFLREVQVNLIDFKKNDY 478
Db 355 ltfndlvkvpclpnpqmmllqpeqlcwisgwgatee-kgktsevinnaakvllietqrcnsr 413
QY 479 LVYDSYLTPRMCAAGDLRGGRDSCQDGGPLVCEQNNRWYLAGVTSWGTGCGQRNPGV 538
Db 414 yvydnlitpamicagflqgnvdsqgdspplvtsnnnllwllgdtswgscakayrpgv 473
QY 539 YTKVTEVLPWIYSKMS 555
Db 474 ygnvmvftdwlyrqmka 490
RESULT 15
AAE06943
ID AAE06943 standard; Protein; 492 AA.
XX
AC AAE06943;
XX
DT 16-OCT-2001 (first entry)
DE Human serine protease protein.
XX
KW Human; transmembrane serine protease; membrane-type serine protease;
KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW serine protease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 255..259 /note= "Conserved motif"
FT
XX WO200157194-A2.
XX PN
XX 09-AUG-2001.
XX PD
XX 02-FEB-2001; 2001WO-US03471.
PF

XX 03-FEB-2000; 2000US-0179982.
PR 18-FEB-2000; 2000US-0183542.
PR 22-JUN-2000; 2000US-0213124.
PR 26-JUL-2000; 2000US-0220970.
PR 08-SEP-2000; 2000US-0657986.
PR 22-SEP-2000; 2000US-0234840.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Ong EO, Yeh J;
XX
XX WPI; 2001-488877/53.
DR N-PSDB; AAD13168.
XX

XX Novel single chain polypeptide comprising protease domain of type-II
PT membrane-type serine protease or its catalytically active portion
PT useful for treating and preventing cancer and tumor
XX

PS Disclosure; Page 252-253; 256pp; English.

XX The invention relates to transmembrane serine proteases and their
CC corresponding nucleotides and the protease domain of a type-II
CC membrane-type serine protease (MTSP). MTSP is useful for identifying
CC compounds that modulate or inhibit its proteolytic activity and for
CC formulating a medicament for treating neoplastic disease. MTSP and
CC its corresponding nucleotides are useful in preventing or treating
CC tumors or cancers such as lung carcinoma, colon adenocarcinoma and
CC ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
CC is useful as a diagnostic marker for tumour development, growth and/or
CC progression and as immunogens to generate antibodies that specifically
CC bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
CC gene therapy. The present sequence is human serine protease protein.
XX

SQ Sequence 492 AA;

Query Match 24.6%; Score 736.5; DB 22; Length 492;
Best Local Similarity 32.0%; Pred. NO. 1.5e-40;
Matches 178; Conservative 86; Mismatches 194; Indels 99; Gaps 18;

QY 27 ACTPPGRA-----SPAQASPAQASPAQTPPGRASPAQASPAQTPPGRASPAQ 78
Db 5 sgspgaigpyyehyqpenypaqbtvpt-vyevhpaqyypspvp-----qyaprv 56
QY 79 ASPARASPALASLSRSSGRSSARSASVTPTRYVLVTRATPVGAVPIKSSPARSAPAT 138
Db 57 ltqa-snpvvtcpkpsg-----tvcstskkalcitltlgtf----- 94
QY 139 RATRESPGTSLPKFTWREGOKQLPLIGCVLLLLIALVVSLLILFQFWQ-----GHTGIRY 192
Db 95 -----lvga-----alaagll-----wkfngskcsngiec 120
QY 193 KEQRESCPKHVRCDGVDCGLKASDELGCVRFDWDKSLKIYSSSHQWLPICSSNWDS 252
Db 121 -dsngtcinpsnwdyghcpgeedenrcvrlvgpnfilqmysqrkshwpcqddwnen 179
QY 253 YSEKTCQQLGFESAHRTEVAHRDFANSFILRYNST-----IOESLHRSE-CPSQRYIS 306
Db 180 ygraacrdmgyknfyssq-givddsgstsfmklntsgnvdlykkllyhsadccskavvs 238
QY 307 LQCSHCGCLR---AMTGRIVGALASDKWPQVSLHFGTHICGGTLIDAQWVLTAAHCF 363
Db 239 lrclacgvnlinsrqrivggesalpgawpqvslhvnqnvhvcggsliitpewivtaahcv 298
QY 364 FVTRKVLGEG---WKVYAGTSLNLHQLPEAA--SIAETIIINSNYTDEDDYDIALMRLSKP 418
Db 299 -----ekplnnpwhwtafagllrqsfmfygagvqvkvishpnysdktkndialmklqkp 354
QY 419 LTLSAHIHPACPLPMHQGTFSINETCWTGKTRTDDTKTSPFLREVQVNLIDFKKCN DY 478
Db 355 ltfdnlvkvclnpgmmldpeqlcwisgwgatee-kgtktsevlnaakvllietgrcnr 413

QY 479 LVTDSYLTPRMWCAGDLRGDRSDSCQDGGPLVCEQNNRWYLAGVTSWGTGCGQRNKPgv 538
Db 414 yyvynlftpamicagflqgnvdsccgdsqgplvtstnnmlwllgdtswsgscakayrpgv 473
QY 539 YTKVTEVLEPWIYSKMES 555
Db 474 ygnvmvftdwiyrqmka 490

Search completed: August 20, 2002, 03:41:36
Job time: 287 sec

Db 239 LRCIACGVNLSSRSQRIYVGGESALPGAMPQVSLHVQNVHVCSSIIIPWIVTAACHV 298
QY 364 FVTRKVLG---WKVYAGTSNHLQPEAA--SIABIIINSNYTDEDDYDIALMRLSKP 418
Db 299 -----EKPLNPNHWTAFAGILROSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKP 354
QY 419 LTLSAHHACPLPMHCQTSNLTCHITGKTRTDDKTSPLREVOVNLIDFKKNDY 478
Db 355 LTFNDLVKPVCLPNCMMIQLPEQLCWISWGATEE-KGKTSEVLNAAKVLLTETQRCNSR 413
QY 479 LVYDSYLTTRMCMAGDLGRGDSQCGSGPLVCEONNRWYLAGVTSWGTGCGQRNKP 538
Db 414 YVYDNLITPAMICAGFLQGVNSQCGSGPLVTSKNNIWLILGDTSWGSGCAKAYRPGV 473
QY 539 YTKVTEVLPWIYSKMESE 556
Db 474 YGNVMVFTDWIYRQMRAD 491

RESULT 2
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/251,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match 24.2%; Score 726; DB 4; Length 454;
Best Local Similarity 37.7%; Pred. No. 4.4e-44;
Matches 154; Conservative 70; Mismatches 162; Indels 22; Gaps 9;

QY 164 IGVLLIALVLSLILFQWOGHTGIRYKEQRESCPKHVRCDGVVDCCKLSKSLGCVR 223
Db 55 LGIIALILALIGLGHFDC----SKYRCRSFKIELITKCDGVSDCKDGEDEYRCVR 110
QY 224 FOWDKSLKIYSGSSHOWLPICSSNMNDSYSEKTCQQLGFESAHRTEVA-----HR 275
Db 111 VGGONAVLQVFTAAS--WKTMCSDDKGHYANVACAQLGFPSPVSSDNLNVSSLEGQFRE 168
QY 276 DFANFSILRYN--STIQSLH-RSCPSQRIISQCSCG-LRAMTGRIVGALASDSK 331
Db 169 EFYSIDHLLPDDKVATLHSHSVYREGCASGHVVTLOCTACGHRGYSRIVSGNMSLLSQ 228
QY 332 WPMQVSLHFTGTHICGTLIDAQWLTAACHFEVTRKYLEGWKVYAGT'SNLHQLPEAAS 391
Db 229 WPMQASLQFGYHLCGGSVTPWLWITAAHCYVDL--YLPKSTIOVLGLVSLDNPAPSH 286
QY 392 IAE-IINSNYTDEDDYDIALMRLSKPITLSAHTPACPLPMHGQTFSLNETCWITGFGK 450
Db 287 LVEKIVYHSKYPKRLGNDIALMKLAGPLTFENMIQVCLPNSEENFPDGKVCWTSGWA 346
QY 451 TRETDDKTSPLREVOVNLIDFKKNDYLVYDSYLTPRMWCAGDLRGDRSCQDGGGL 510
Db 347 TEDGGD-ASPLNHAAVPLISNKRICHNRDVGIGTISPSMLCAGYLTGGVDSCQDGGGL 405
QY 511 VCBQNNRWLAGVTSWGTGCGQRNKPQVTKVTEVLPWIYSKMESEVR 558

Db 406 VCQERRLWLKVGATSEGICAEVKNKPGVYTRVTSFLDWIHEQMERDLK 453
RESULT 3
US-09-008-271A-6
; Sequence 6, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0355
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match 23.2%; Score 696.5; DB 4; Length 435;
Best Local Similarity 34.9%; Pred. No. 5.2e-42;
Matches 152; Conservative 73; Mismatches 163; Indels 47; Gaps 11;

QY 149 LPKFTWREGOKLPLGICVLLIALVLSLIL-----FQFWQGHGTGIRYKEQRESCPKH 202
Db 22 IPMEYFR--KVGPIIALLSLASIIIVVVLKIVLKDYYFLCGQ-----PLH 67
QY 203 AV-----RCGVVVDCKLKSDELGC-----VRFMDKSLKLIYSSSSHOWLPICSS 247
Db 68 FIPRKQLCDGELDCPLGEDDEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFD 127
QY 248 NNWDSYSEKTCQQLGFES--AHRTEVAHRDFANSFILRYNSTIOESLHRSECPQRYI 305
Db 128 NTEALAEATACOMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGFCLSGLV 187
QY 306 SLQCSHCHGLRAMTGRIVGGALASDSKWPQVSLHFGTTHICGTLIDAQWLTAACHFEV 365

```

Db 188 SLHCLACGESLTPRVVGGESVSDSWPQVSTQYDKQHVCGSILDPHWLTAACHF-- 245
QY 366 TREKVLGHKVTAGTSLNHLQPEASIAIILII--NSNTYDEDDYDIALMRLSKPLTILS 422
Db 246 RKHTDVFNNKVRAGSDKLGFFSLA-VAKIIITEFNPMY---PKNDIALMKLOFPLTFS 301
QY 423 AHIHPACLPMHGQTFSLNETCWTGKTRTDDKTSPLREYQVNLIDFKKCNLYVD 482
Db 302 GTVPRICLPFDEELTPATPLHIGWFTKNGGRMSDILLQASVOVIDSTRCNADDAVQ 361
QY 483 SYLTPMCMAGDLRGRSDCGSDGSGPLVCEQNNRWYLAGVTSWGTGCGRKNPGVYTKV 542
Db 362 GEVTEKMMCAGIEGGVDTCQDGGGLM-YQSDQWVVIVSWGYCGGPGSTPGVYTKV 420
QY 543 TEVLPMYISKMESEV 557
Db 421 SAYLNIYVNWRAEL 435

RESULT 4
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TAGD-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2

Query Match 21.3%; Score 637.5; DB 4; Length 455;
Best Local Similarity 35.5%; Pred. No. 8.4e-38;
Matches 145; Conservative 68; Mismatches 173; Indels 23; Gaps 10;

QY 164 IGVLLIALVSLILFQWOGHTGIRYKEQRESCPKHVRCDGVVDCKLKSDLGCVR 223
Db 55 LGIIALILALAIGLGHFDC---SGKYRCRSFKEIITRCDGVSCKDGEDEYRCVR 110
QY 224 FDMDSLLKIYSGSSHOWLPICSSNWNDSYSEKTCOOLGFESAHRTTEVA-----HR 275
Db 111 VGGQNAVLOVFTAAS--WKTMCSDMWKGHYANVACAQLGFPSTVSDNLVRVSSLEGQFRE 168
QY 276 DFANFSILRYN--STIQESLH-RSECPQRYISLQCSHCG-LRAMTGRIVGGALASDSK 331
Db 169 EFPVSDHLLPDDKVTALHHSVVRECCASGHVVTLOCTACGHRGYSRIVGVGNMSSLQ 228
QY 332 WPMQVSLHGTTHICGGTLIDAQWLTAACHCFVTRKVLGKVKYVAGTSLNHLQPEAAS 391
Db 229 WPMQASLOPQGHLCGGSVITPLWIIITAACHCVYDL--YLPKSWTIQVGLVSLIDNPAPSH 286
QY 392 IAE-IILNSNYTDEDDYDIALMRLSKPLTSLAHHPACLPMHGQTFSLNETCWTGKTR 450
Db 287 LVEKIVYHSKPKRLGNIDIALMKLAGPLTFNEMIQPVCLPNSEENFPDGKVCWTSGWA 346
QY 451 TRETDDKTSPLREYQVNLIDFKKCNLYVDYSLTPMCMACAGDLRGR--DSCQGDGSGGP 509
Db 347 TEDGGD-ASPVNLHAAVPLISKNKDLQPGQVRWVHLLPLHALRGLPDGWRWNSCQGDGSGGP 405
QY 510 LVCEQNNRWYLAGVTSWGTGCGQRNKPQGYTKVTEVLPWYISKMESEVR 558
```

```

Db 406 LVQERRLKLKVGATSGFICGADVKNPKVGYTRVTSFLDWIHEQMERDLK 454

RESULT 5
US-09-000-846-2
; Sequence 2, Application US/09000846
; Patent No. 5981830
; GENERAL INFORMATION:
; APPLICANT: WU, QINGYU
; APPLICANT: SADLER, JASPER
; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
; TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,846
; FILING DATE: 30-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION NUMBER: US 08/866,058
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: LEBOWITZ, RICHARD M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: BERLX 65P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-000-846-2

Query Match 20.7%; Score 619.5; DB 2; Length 416;
Best Local Similarity 37.1%; Pred. No. 1.4e-36;
Matches 141; Conservative 61; Mismatches 133; Indels 45; Gaps 12;

QY 209 VYDCKLKSD--LGCVRFDWDKSLKIYSGSSHOWLPICSSNWNDSYSEKTCOOLGFESA 266
Db 38 IYVILQSDQEPYQVQLSPGDSRLAVFDKTEGTWRLCLSSRNARVAGLGCCEMGLRA 97
QY 267 HRTTEVAHRDF-----ANSFS-----ILRYNSTIOESLHRSCEPQRYISLQCSHC 312
Db 98 -----LAHSELDRVTRAGANTSGFFCVDEGGLRLAORLIDVSVCDPGRFRITATCQDC 152
QY 313 GLRAM-TGRIVGGALASDKWPMQVSLHGTTHICGGTLIDAQWLTAACHCFVTRKVL 371
Db 153 GRKLPVDRIVGGQSDSSLGRWPQVSLRYDGTGHLCCGSLGSDWLTAAHC-FPERNRVL 211
QY 372 EGVKVTAG-----TSNLHQLPEAA--SIAELIINSNY-----TDEDDYDIALMRLSKPL 419
Db 212 SWRVFVAGAVARTS-----PHAVQLGVQAVIYHGGVLPFRDPTIDENSNDIALVHLSL 266
QY 420 TSLAHHPACLPMHGQTFSLNETCWTGKTRTDDKTSPLREYQVNLIDFKKCNLY 479
Db 267 PLTEYIQPVCLPAAGQALVDGKVTVTWGNT-QFYGQAMVLOEARVPIISNEVCNSPD 325
```

QY 480 VYDSYLTPTMCMACADLRGGRDSCQSGGGLVCEQN-----NRWYLAGVTSWGTGCGQRNK 535
Db 326 FYGNQIKPKMFCAGYPEGIDACQDGGSPFVCEDSISGTSRWRLCGIVSWGTCALARK 385
QY 536 PGVYTKVTEVLPWPIYSKMES 555
Db 386 PGVYTKVTDREWIFRAIKT 405

RESULT 6
US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO1
; CLONE: 556016
; US-08-807-151-1

Query Match 19.7%; Score 592; DB 3; Length 283;
Best Local Similarity 42.7%; Pred. No. 8e-35;
Matches 119; Conservative 44; Mismatches 102; Indels 14; Gaps 6;
QY 287 NSTIOESLRHSE-CPQSRYISLQCSHCLGR---AMTGRIVGGALASDSKWPQVSLHFGT 342
Db 9 NVDIYKKLTHSDACSKAVVSLRCLACGVNLNLSRSRISRVGGESALPGAWPQVSLHVON 68
QY 343 THICGGTLLDAQWVLTAAHCFVTRKVLLEG---WKVYAGTSLNLHOLPEAA--STAEII 397
Db 69 VHVCGSIIPTPEWYTAACHV---EKPLNNPWHWTAFAGILRQSFMEYAGYQVEKVL 124
QY 398 NSNYTDEDDYDIALMRLSKPLTSLAHIIHACPLPHMGQFTSLNETCWITGFKTRETDDK 457
Db 125 HPNYSKTKNNIDIALMKLQKPLTFLNDLVRPCLPDPGMMLOPEQLCWISGWGATEE-KGK 183

QY 458 TSFFLEREVQNLIDFKKNDYLVYDSYLTPTMCMACADLRGGRDSCQSGGGLVCEQN 517
Db 184 TSEVLNAAKVLTIETQCNRSRYVDNLITPAMICAGFLQGNVDSCQSGGGLVTSKNNI 243
QY 518 WYLAGVTSWGTGCGQRNKPQVYTKVTEVLPWPIYSKMESE 556
Db 244 WMLIGDTSWGSCKAYRPGVYGNVMVFTDWIYRQMRAD 282
RESULT 7
US-09-478-957-1
; Sequence 1, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO1
; CLONE: 556016
; US-09-478-957-1

Query Match 19.7%; Score 592; DB 4; Length 283;
Best Local Similarity 42.7%; Pred. No. 8e-35;
Matches 119; Conservative 44; Mismatches 102; Indels 14; Gaps 6;
QY 287 NSTIOESLRHSE-CPQSRYISLQCSHCLGR---AMTGRIVGGALASDSKWPQVSLHFGT 342
Db 9 NVDIYKKLTHSDACSKAVVSLRCLACGVNLNLSRSRISRVGGESALPGAWPQVSLHVON 68
QY 343 THICGGTLLDAQWVLTAAHCFVTRKVLLEG---WKVYAGTSLNLHOLPEAA--STAEII 397
Db 69 VHVCGSIIPTPEWYTAACHV---EKPLNNPWHWTAFAGILRQSFMEYAGYQVEKVL 124
QY 398 NSNYTDEDDYDIALMRLSKPLTSLAHIIHACPLPHMGQFTSLNETCWITGFKTRETDDK 457
Db 125 HPNYSKTKNNIDIALMKLQKPLTFLNDLVRPCLPDPGMMLOPEQLCWISGWGATEE-KGK 183

```
QY 458 TSPFLREVQVNLIDFKCNDYLYDSYLTTPMMACAGDLRGRDSCQDGGPLVCEQNNR 517
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 184 TSEYNAAKLLLETTORCNRGRYVYDNLITPAMICAGFLOGNVDSCQDGGXLVTSKNNI 243
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 518 WYLAGVTSWGTGCGQRNKPQVYTKVTEVLPWYISKMESE 556
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 244 WLLIGTWSGCGCAKAPRGVYGNVYFTDWIRQMRAD 282
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 8
US-08-681-151-3
; Sequence 3, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael
; APPLICANT: Goli, Surya
; TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,151
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0074US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 205011
US-08-681-151-3

Query Match 18.9%; Score 565.5; DB 2; Length 638;
Best Local Similarity 33.5%; Pred. No. 1.7e-32;
Matches 130; Conservative 56; Mismatches 95; Indels 107; Gaps 15;

QY 196 RESCPKHAVRCGDW-----DCKLKSDELGC---VRFDWDKSLTKT---YSGSSHOWLP 243
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 319 QETCTK-TIRCOFTYSLLPQDCKAE---GCKSLRLSDGSPFRITTYEAQSS----- 368
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 244 ICSSNNDYSSEKTCQOLGFESAHRTTEVAHRDFANSFSLRYNSTIQESLHRSECPQR 303
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 369 -----GYSLRLCKVW--ESSDCTKI----- 387
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 304 YISLQSHCGLRAMTGRIVGGALASDSKWPQVSLH---FGTHICGGTLIDAQWLTAA 360
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
```

```
Db 388 -----NARIVGGTSSSLGWPQVSLQVKLVSONHMCQSGSIIGROWILTAA 433
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 361 HCFVTREREKVLG-----WKVYAGTSLNHLPEA---ASTAEIIINSNYTDEDDYDIA 411
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 434 HCF-----DGIPYDPVWRIYGIILNLSITNKTTFESSIKELIIHQYKMSGSDYDIA 485
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 412 LMRSLKPLTSLSAIHAPACLPMHGQTFSLNTECWTGTGKTRETDKTSFPLREVQVNLID 471
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 486 LIKLQTLPLNTEFQKPICLPSKADNTIYTNWVTGWTGTYKERGE-TQNIILQKATIPLPV 544
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 472 FKKC-----NDYLYDSYLTTPMMACAGDLRGRDSCQDGGPLVCEQNNRWYLAGVTSW 527
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 545 NEECQKKRYDYVI-----TKOMICAGYKEGIDACKDGSGLVCKHSGRWQLVGIITSW 599
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 528 TGCQRNKPQVYTKVTEVLPWYISKMES 555
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 500 EGCAKEQPGVYTKVAEYIDWILEKIQS 627
|| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 9
US-08-944-483-63
; Sequence 63, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-944-483-63
```



```
QY 417 KPLTSAHIHPACLP MHGOTFSLNETCWTGKTRTDDKTSPLFREVQVNLIDPKKN 476
Db 663 MKVNTDYIQCIPLEENOVFPGRICSTAGWG-ALIIYQGSTADVLQEADVPLLSNEKQ 721
QY 477 DYLVVDSYLTPRMKAGDLRGDRSCQDGGPLVCEQNNRYLAGVTSWGTGCGORNP 536
Db 722 QOMP-EYNITENNVAGYBAGVDCQDGGPLMCQENNRLLAGVTSFGYOCALPNRP 780
QY 537 GYTKVTEVLPWIS 551
Db 781 GYVARVPRTEWIS 795

RESULT 12
US-09-027-337-3
; Sequence 3, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hiroto
; TITLE OF INVENTION: TAGG-15: An Extracellular Serine Protease Overexpressed in
; FILE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 3
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in TAGG-15
US-09-027-337-3

Query Match 17.7%; Score 530.5; DB 2; Length 256;
Best Local Similarity 42.6%; Pred. No. 1.6e-30;
Matches 106; Conservative 38; Mismatches 88; Indels 17; Gaps 6;

QY 320 RIVGALASDSKWPQVSLHFGTTHICGGTLIDAQWVLTAAHCFVTRKVLGKVVYAG 379
Db 1 RIVGGRDTSLGRRPQVSLRYDGAHLGGSLSGDWVLTAAHC-FPERNRVLSRWRFAG 59
QY 380 T---SNLHQLPEAASIAEIIINSNY-----TDEDDYDIALMRLSKPLTLSAHIHPACL 430
Db 60 AVAQAASPHGL--QLGVQAVVYHGGYLPFRDPNSENNDIALVHLSSPLPLTEYIQPVCL 117
QY 431 PMHGOTFSLNETCWTGKTRTDDKTSPLFREVQVNLIDFKKNDYLVYDSYLTPRMK 490
Db 118 PAAGQALVDGKICTVTGWTGNTQYQQAG-VLQEARVPIISNDVNCGADFYGNQIKPMF 176
QY 491 CAGDLRGDRSCQDGGPLVCE---QNNRWYLAGVTSWGTGCGORNPVYTKVTEVL 546
Db 177 CAGYEGGIDACQDGGPGFVCEDSISRTPRWRLCGIVSWGTCALAQKPGVYTKVSDPR 236
QY 547 PWYSKMS 555
Db 237 EWIFAOKT 245

RESULT 13
US-09-944-483-67
; Sequence 67, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
```

```
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-67

Query Match 17.5%; Score 525.5; DB 4; Length 255;
Best Local Similarity 42.3%; Pred. No. 3.7e-30;
Matches 105; Conservative 38; Mismatches 88; Indels 17; Gaps 6;

QY 321 IVGGALASDSKWPQVSLHFGTTHICGGTLIDAQWVLTAAHCFVTRKVLGKVVYAGT 380
Db 1 IVGGRDTSLGRRPQVSLRYDGAHLGGSLSGDWVLTAAHC-FPERNRVLSRWRFAGA 59
QY 381 ---SNLHQLPEAASIAEIIINSNY-----TDEDDYDIALMRLSKPLTLSAHIHPACL 431
Db 60 VAQAASPHGL--QLGVQAVVYHGGYLPFRDPNSENNDIALVHLSSPLPLTEYIQPVCL 117
QY 432 MHGOTFSLNETCWTGKTRTDDKTSPLFREVQVNLIDFKKNDYLVYDSYLTPRMMC 491
Db 118 AAGQALVDGKICTVTGWTGNTQYQQAG-VLQEARVPIISNDVNCGADFYGNQIKPMFC 176
QY 492 AGDLRGDRSCQDGGPLVCE---QNNRWYLAGVTSWGTGCGORNPVYTKVTEVLP 547
Db 177 AGYEGGIDACQDGGPGFVCEDSISRTPRWRLCGIVSWGTCALAQKPGVYTKVSDFRE 236
QY 548 WIYSKMS 555
Db 237 WIFAOKT 244

RESULT 14
US-08-944-483-64
; Sequence 64, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
```

Search time (sec): 79.410000

[illegible]

532 lnArgAsnLysProGlyValThrLysValThrGluValLeuProTrp 548
:::|||||
1400 AAGCTTACAGACGAGAGTGTACGGGAATGTGATGATTTCAGGACTGG 1449

549 ileTyrSerLysMetGluSerGlu 556
|||||
1450 ATTATCGACAATGAGGCAGAC 1473

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-342-749-29

seq_documentation_block:

; Sequence 29, Application US/09342749

; Patent No. 6166194

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Teng, David H.-F.

; APPLICANT: Myriad Genetics, Inc.

; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor

; FILE REFERENCE: 2318-202

; CURRENT APPLICATION NUMBER: US/09/342,749

; CURRENT FILING DATE: 1999-06-29

; EARLIER APPLICATION NUMBER: US 60/091,044

; EARLIER FILING DATE: 1998-06-29

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 29

; LENGTH: 2479

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-342-749-29

alignment_scores:

Quality: 736.50 Length: 557

Ratio: 2.239 Gaps: 18

Percent Similarity: 59.066 Percent Identity: 31.957

alignment_block:

US-09-879-792-12 x US-09-342-749-29 ..

Align seg 1/1 to: US-09-342-749-29 from: 1 to: 2479

27 AlaGlyThrProGlyArgAla.....Se 35

:::|||||
69 TCAGGTCAACCAAGTATTGGACCTTACTATGAACCAATGATGATACCA 118

35 rProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrProG 52

:::|||||
119 ACCGGAACCCCTATCCGCACAGCCCACTGTGTCCCTCCACT...GTCT 165

52 lyArgAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArg 68

:::|||||
166 ACGAGTGCATCCGCTCAGTACTACCGCTCCCGGTGCC... 206

69 AlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaAlaSe 85

:::|||||
207CAGTACGCCCGGAGGTCTCGACGAGGCT...TCCAA 241

85 rProAlaLeuAlaSerLeuSerArgSerSerSerGlyArgSerSerA 102

:::|||||
242 CCCGTCGTCGTGCAGCGACGCCCAATCCCATCCGCGG..... 278

102 laArgSerAlaSerValThrThrSerProThrArgValThrLeuVal 118

:::|||||
279ACAGTGGACCTCAAGACTAAGAAAGCACTGTGCATC 317

119 AlaThrProValGlyAlaValProIleArgSerSerProAlaArgSer 135

:::|||||
318 ACCTTGACCCCTGGGACCTTC..... 338

135 aProAlaThrArgAlaThrArgGluSerProGlyThrSerLeuProLys 152

338 338
152 heThrTrpArgGluGlyGlnLysGlnLeuProLeuIleGlyCysValLeu 168
:::|||||
339CTCGTGGAGCT..... 350
169 LeuLeuIleAlaLeuValSerLeuIleLeuPheGlnPheTrpG1 185
|||||
351GCCTGGCGCTGGCTACTC.....TGGAA 376
185 n.....GlyHisThrGlyIleArgTyrLysGluGlnA 196
:::|||||
377 GTTCATGGGCAGCAAGTCTCCAACTCTGGGATAGAGTGC...GACTCCT 423
196 rgGluSerCysProLysHisAlaValArgCysAspGlyValValAspCys 212
:::|||||
424 CAGGTACCTGCATCAACCCCTCTAAGTGGTGTGATGGGTGCACACTGC 473
213 LysLeuLysSerAspGluLeuGlyCysValArgPheAspTrpAspLys 229
:::|||||
474 CCCGGCGGGAGGAGAGAACTGGTGTCTTCGCCCTACGGACCAACTT 523
229 rLeuLeuLysIleTyrSerGlySerHisGlnTrpLeuProIleCysS 246
:::|||||
524 CATCCTTCAGATGTACTCTCAGAGGAAGTCTTGGCACCTGTGTGCC 573
246 erSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGlnLeuGly 262
:::|||||
574 AAGCAGCTGGAAACGAGAACTACGGCGGGCGGCTGCAGGGACATGGGC 623
263 PheGluSerAlaHisArgThrThrGluValAlaHisArgAspPheAlaAs 279
:::|||||
624 TATAAGATAATTTTACTTAGCCAA...GGAATAGTGGATGACAGCGG 670
279 nSerPheSerIleLeuArgTyrAsnSerThr.....IleG 291
:::|||||
671 ATCCACAGCTTATGAACACTGAACACAGTCCCGCAATGTCGATATCT 720
291 lnGluSerLeuHisArgSerGlu...CysProSerGlnArgTyrIleSer 306
:::|||||
721 ATAAAAAAGTGTACCAAGTGTGCTTCTTCAAAAGCAGTGGTTCT 770
307 LeuGlnCysSerHisCysGlyLeuArg.....AlaMetThrGlyArg 320
:::|||||
771 TTACGCTGTTTAGCCTCGGGGTCAACTTGAACCTCAAGCCGCGCAGAGCAG 820
320 gIleValGlyGlyAlaLeuAlaSerAspSerLysTrpProTrpGlnValS 337
:::|||||
821 GATCGTGGCGGTGAGACGCGCTCCCGGGGGGCTTGGCCCTGGCAGGTCA 870
337 erLeuHisPheGlyThrThrHisIleCysGlyGlyThrLeuIleAspAla 353
:::|||||
871 GCCGTCAGCTCCAGACGCTCCAGTGTCCGAGGCTCCATCATCATCCCC 920
354 GlnTrpValLeuThrAlaAlaHisCysPheValThrArgGluLysVa 370
:::|||||
921 GAGTGGATCGTGACAGCGCCCACTGCTG.....GAAAAACC 958
370 lleGluGly.....TrpLysValTyrAlaGlyThrSerAsnLeuH 384
:::|||||
959 TCTTAACAATCCATGGCATTTGGACGGCATTTGCGGGGATTTTGAGACAAT 1008
384 lsGlnLeuProGluAlaAla.....SerIleAlaGluIleIleAsn 398
:::|||||
1009 CTTTCATGTTCTATGGACCCCGATACCAAGTACAAAAAGTATTCTCAT 1058
399 SerAsnTyrThrAspGluAlaAspTyrAspIleAlaLeuMetArgLe 415
|||||
1059 CCAAAATTATGACTCAAGACCAAGAACAAATGACATTGCGCTGATGAAGCT 1108
415 uSerLysProLeuThrLeuSerAlaHisIleHisProAlaCysLeuProm 432
:::|||||

```

1109 GCAGAGCCTCTGACTTTCACAGCACCTAGTCAAAACCACTGTGTCTGCCCA 1158
432 eHisGlyInThrPheSerLeuAsnGluThrCysTrpIleThrGlyPhe 448
||||| : : : : : ||||| : : : : :
1159 ACCCAGCATGATCTGCAGCAGAACAGCTCTCTGGATTTCGGGTGG 1208
449 GlyLysThrArgGluThrAspLysThrSerProPheLeuArgGluVa 465
||||| : : : : : ||||| : : : : :
1209 GGGGCCACCGAGGAG...AAGGGAAGACCTCAGAGTGCTGAACGCTGC 1255
465 lGlnValAsnLeuIleAspPheLysLysCysAsnAspTyrLeuValTyrA 482
||||| : : : : : ||||| : : : : :
1256 CAAGTGCTCTCATTTGAGACACAGATGCAACAGCAGATATGCTATG 1305
482 spSerTyrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyCly 498
||||| : : : : : ||||| : : : : :
1306 ACAACCTGATCACACCAGCAGATGATCTGTGCGGCTCTCCAGGGGAAC 1355
499 ArgAspSerCysGlnGlyAspSerGlyClyProLeuValCysGluGlnAs 515
||||| : : : : : ||||| : : : : :
1356 GTCGATCTTCCAGGGTGACAGTGGAGGGCTCTGGTCACTTCGAACAA 1405
515 nAsnArgTyrPyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyG 532
||||| : : : : : ||||| : : : : :
1406 CAATATCTGGTGGCTGATAGGGGATACAAAGCTGGGTCTCGCTGGCCA 1455
532 lNArgAsnLysProGlyValTyrThrLysValThrGluValLeuProTrp 548
::: : : : : ||||| : : : : :
1456 AAGCTTACAGCAGGAGGTGTACGGGAATGTGATGTTATTCACGAGCTGG 1505
549 lIleTyrSerLysMetGluSer 555
||||| : : : : :
1506 ATTTATCGACAAATGAAGCCA 1526

```

seq_name: /cgn2_6/ptodata/2/ina/6B_comb.seq;us-09-518-046-1

```

seq_documentation block:
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192C1P
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
; US-09-518-046-1

```

```

alignment_scores:
  Quality: 727.50      Length: 501
  Ratio: 2.324         Gaps: 13
  Percent Similarity: 62.475  Percent Identity: 33.932

```

alignment_block:

US-09-879-792-12 x US-09-518-046-1 ..

Align seg 1/1 to: US-09-518-046-1 from: 1 to: 2413

```

107 valThrThrSerProThrArgValTyrLeuValArgAlaThrProValG1 123
||||| : : : : : ||||| : : : : :
30 GTAACACTGTGGCCTACTATCTCTCCGTGGTGCCATCATATTTTGGG 79

```

```

123 y.....AlaValProIleArgSerSerProAlaArgSerAlap 136
| : : : : ||||| : : : : : ||||| : : : : :
80 ACTCGGGAATTATGAGGTAGAGTGGAGCGGAGCCGGATGTCAGAGGTC 129
| : : : : ||||| : : : : : ||||| : : : : :
136 ro.....AlaThrArgAlaThrArgGluSerProGlyThrSerLeuPro 150
|| : : : : ||||| : : : : : ||||| : : : : :
130 CTGAATAGTCACCATGGGGAATAATGATCCGCCTGCTGTGTAAGCCCCC 179
| : : : : ||||| : : : : : ||||| : : : : :
151 LysPheThrTrpArg..... 155
| : : : : ||||| : : : : : ||||| : : : : :
180 ..TTCTCATTCGCATCGCTTTTGGCCTTGTATGATTGAAATAAAGTCC 226
| : : : : ||||| : : : : : ||||| : : : : :
155 ..... 155
227 TGTTCACACAGATCCAGATGCTGTGTGTGCACACATCTGTCACTGCTGC 276
156 ....GluGlyGlnLysGlnLeuProLeuIleGlyCysValLeuLeuLeu 170
|| : : : : ||||| : : : : : ||||| : : : : :
277 CATTTGAAGTTTTTTTCCCAATCATCGTCATTGGGGATCATTTGCATTGATA 326
| : : : : ||||| : : : : : ||||| : : : : :
171 lIleAlaLeuValValSerLeuIleIleLeuPheGlnPheTrpGlnGlyHi 187
::: : : : : ||||| : : : : : ||||| : : : : :
327 TTAGCACTGGCCATTGTGTGGGCATCCACTTCGACTGC..... 365
187 sThrGlyIleArgTyrLysGluGlnArgGluSerCysProLysHisAlav 204
::: : : : : ||||| : : : : : ||||| : : : : :
366 .TCAGGGAAGTACAGATGCTCGCTCATCTCTTTAAGTGTATCGAGCTGTAA 414
| : : : : ||||| : : : : : ||||| : : : : :
204 alArgCysAspGlyValValAspCysLysLysLeuLysSerAspGluLeuGly 220
|| : : : : ||||| : : : : : ||||| : : : : :
415 CTCGATGTGCGGAGTCTCGGATTGCAAGACGGGGAGGACGAGTACCGC 464
| : : : : ||||| : : : : : ||||| : : : : :
221 CysValArgPheAspTrpAspLysSerLeuLysIleTyrSerGlySe 237
|| : : : : ||||| : : : : : ||||| : : : : :
465 TGTGTCCGGGTGGGTGCTCAGAAATGCGGTGCTCCAGGTGTTTCACAGCTGC 514
| : : : : ||||| : : : : : ||||| : : : : :
237 rSerHisGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrS 254
::: : : : : ||||| : : : : : ||||| : : : : :
515 TTCG.....TGGAAGACCATGTGCTCGATGACTGGAAGGTCACCTACG 558
| : : : : ||||| : : : : : ||||| : : : : :
254 erGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThr 270
|| : : : : ||||| : : : : : ||||| : : : : :
559 CAATATGTTCCCTGTGCCCAACTGGGTTTCCCAAGCTATGTGAGTTCAGAT 608
| : : : : ||||| : : : : : ||||| : : : : :
271 GluValAla.....HisArgaspPheAlaAs 279
| : : : : ||||| : : : : : ||||| : : : : :
609 AACCTCAGAGTGAAGTCCGTCGAGGGGCGAGTTCGGGGAGGAGTTGTGTC 658
| : : : : ||||| : : : : : ||||| : : : : :
279 nSerPheSerIleLeuArgTyrAsn.....SerThrIleGlnGluSerL 294
| : : : : ||||| : : : : : ||||| : : : : :
659 CATCGATCACCTCTTCCAGATGACAGAGTGCATTCATACACCACCTCAG 708
| : : : : ||||| : : : : : ||||| : : : : :
294 euHis...ArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCys 309
|| : : : : ||||| : : : : : ||||| : : : : :
709 TATATGTGAGGGAGGATGTCCCTCTGGCCACGCTGGTTACCTTCAGTGC 758
| : : : : ||||| : : : : : ||||| : : : : :
310 SerHisCysGly...LeuArgAlaMetThrGlyArgIleValGlyGlyAl 325
::: : : : : ||||| : : : : : ||||| : : : : :
759 ACAGCTGTGTGTCATAGAAGGGGCTACAGCTCAGCGATCGGTGGGTGAAA 808
| : : : : ||||| : : : : : ||||| : : : : :
325 aLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyT 342
::: : : : : ||||| : : : : : ||||| : : : : :
809 CATGTCCCTGTGCTCGCAGTGGCCCTGGCAGGCCAGCCCTTCAGTTCAGG 858
| : : : : ||||| : : : : : ||||| : : : : :
342 hrThrHisIleCysGlyGlyThrLeuIleLeuAspAlaGlnTrpValLeuThr 358
|| : : : : ||||| : : : : : ||||| : : : : :
859 GCTACCACCTGTGCGGGGCTGTGTCATCAGCCGCCCTGTGGATCATCACT 908
| : : : : ||||| : : : : : ||||| : : : : :
359 AlaAlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLy 375
||||| : : : : : ||||| : : : : :
909 GCTGCACACTGTGTTTATGACTTG.....TACCTTCCCAAGTCATGAC 952
| : : : : ||||| : : : : : ||||| : : : : :

```



```
759 ACAGCCTGTGTCATAGAGGGCTACAGCTCAGCATCTGGTGGGAAA 808
325 aleuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyT 342
      ::::: |||||:::|||||:::|||||:::|||||:::|||||
809 CATGTCCTGTCTCGCAGTGGCCCTGCGAGCCAGCCCTCAGTTCCAGG 858
342 hrThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThr 358
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
859 GCTACCACTGTGGGGGGCTGTGTCATCAGCGCCCTGTGATCATCACT 908
359 AlaAlaHisCysPhePhe.....ValThrArgGluLysVal..... 370
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
909 GCTGCACACTGTGTTATGAGATTGTAGCTCCTAGAGAAAGGCGACAG 958
371 .....:::|||||:::|||||:::|||||:::|||||:::|||||
959 AAGAGGAAGAGCTCCTGTGCTGGAGGAAACCCACAAAAAATGAAGGAC 1008
374 .....:::|||||:::|||||:::|||||:::|||||:::|||||
1009 CTAGACCTTCCATAGCTAATTCAGTGGACCATGTTATGGCAGATACAG 1058
378 AlaGlyThrSerAsnLeuHis.GlnLeuProGluAlaAlaSerIle.... 392
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1059 CTTGTACTCTCCCAAGTCATGGACCATCCAGGTGGTCTAGTTTCCCTG 1108
393 .....:::|||||:::|||||:::|||||:::|||||:::|||||
1109 TTGGACAATCAGCCCATCCACTTGGTGGAGAAAGATTGTCTACCACAG 1158
399 rAsnTyrThrAspGluAspAspTyrAspIleAlaLeuMetArgLeuS 416
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1159 CAAGTACAAGCAAGAGGCTGGCAATGACATCGCCCTTATGAAGCTGG 1208
416 eRLysProLeuThrSerAlaHisIleHisProAlaCysLeuProMet 432
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1209 CCGGGCCACTCAGCTTCAATGAATGATCCAGCTGTGTGCTGCCCAAC 1258
433 HisGlyGlnThrPheSerLeuAsnGluThrCysTrpIleThrGlyPheG 449
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1259 TCTGAAGAGAACTTCCCGGATGGAAGAGTGTGTGGACGTGAGGATGGG 1308
449 yLysThrArgGluThrAspAspLysThrSerProPheLeuArgGluValG 466
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1309 GGCCACAGAGGATGGAGGTGAC...GCCCTCCCTGTCTGAACCACGCGG 1355
466 lnValAsnLeuIleAspPheLysCysAsnAspTyrLeuValTyrAsp 482
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1356 CCGTCCCTTTGATTCCACACAGATCTGCACACACAGGGACGTGTACGGT 1405
483 SerTyrLeuThrProArgMetMetCysAlaGlyAspLeuArgGlyGlyAr 499
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1406 GGCATCATCTCCCTCCATGCTCTGCGGGGCTACCTGACGGGTGCGGT 1455
499 gAspSerCysGlnGlyAspSerGlyGlyProLeuValCysGluClnAsnA 516
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1456 GGACAGCTGCCAGGGGACAGCGGGGGGCCCTGTGTGTCAAGAGAGGA 1505
516 snArgTrpTyrLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGln 532
      :::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1506 GGCTGTGGAAGTTAGTGGAGCGACCAAGCTTGGCATCGCTGGCGACAG 1555
533 ArgAsnLysProGlyValTyrThrLysValThrGluValLeuProTrpIl 549
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1556 GTGAACAGGCTGGGGGTGTACACCCGCTACCCCTTCCCTGGACTGGAT 1605
549 eTyrSerLysMetGluSerGluValArg 558
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1606 CCACGACAGATGGAGAGACCTTAAAA 1633
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-09-008-271A-18

seq_documentation_block:

^

```
; Sequence 18, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
;
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008, 271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
;
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18
```

alignment_scores:

Quality:	707.50	Length:	530
Ratio:	2.211	Gaps:	16
Percent Similarity:	60.377	Percent Identity:	32.642

alignment_block:

US-09-879-792-12 x US-09-008-271A-18 ..

Align seg 1/1 to: US-09-008-271A-18 from: 1 to: 2038

58	GlnAlaSerProAla...	GlyThrProGlyArgAlaSerProGlyAr	73
:::			
42	AAGCGCTGCCCTGCAC	TGCGGGCCCTCCAGCCAGTGTGACCGGGACT	91
73	gAlaSer.....	ProAlaGlnAlaSerProAlaArgAlaSerProAlaL	88
:::		:::	
92	TCTACCTGTGGCCAGC	AGCTGTGTGGGAGGCCCTCTCTGCTGCC	141
88	euAlaSerLeuSerArg	SerSerSerGlyArgSerSerAlaArgSer	104
	:::		:::
142	TT...GGGGTGACAAT	CTCAGCTCCAGGTACAGGAGACCGGGAGGATC	188
105	AlaSerValThrThrSer	ProThrArgValTyrLeuValArgAlaThrPr	121

189 ACAGAGCCAGCATGGATCCTGACAGTGTCAACCTCTGAACAGCCTCG... 236
121 oValGlyAlaValProIleArgSerProAlaArgSerAlaProAlaT 138
237ATGCAAAACCCCTGGCAAC..... 257
138 hrArgAlaThrArgGluSerProGlyThrSerLeuProLysPheThrTrp 154
258CCGGTATCCCATGG 272
155 ArgGluGlyGlnLys..GlnLeuProLeuIleGlyCysValLeuLeuLeu 170
273 AGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGCGCTG 322
171 IleAlaLeuValValSerLeuIleLeu.....Ph 181
323 CGGAGTATCATATGTTGGTGTCTCATCAAGGTGATTCGTGATAAATA 372
181 eGlnPheTrpGlnGlyHisThrGlyIleArgTyrLysGluGlnArgGluS 198
373 CTACTTCTCTGGGGCAG..... 391
198 erCysProLysHisAlaVal.....ArgCysAspGlyValVal 210
392CTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGAGCTG 436
211 AspCysLysLeuLysSerAspGluLeuGlyCys..... 221
437 GACTGTCCCTGGGGAGGAGGAGGACCTGTGTCAAGAGCTTCCCGCA 486
222ValArgPheAspTrpAspLysSerLeuLeuLysI 233
487 AGGCGCTGCAGTGGCAGCTCCGCTCTCCAGGACCGATCCACATGCAGG 536
233 leYrSerGlySerSerHisGlnTrpLeuProIleCysSerSerAsnTrp 249
537 TGTGGACTCGGCCACAGGAACCTGTTCTGCTCTTTCGACAACTTC 586
250 AsnAspSerTyrSerGluLysThrCysGlnGlnLeuGlyPheGluSer.. 265
587 ACAGAAGCTCTCGCTGAGACAGCTGTAGCGAGATGGCTACAGCACAA 636
266AlaHisArgThrThrGluValAlaHisArgAspPheAlaAsnSerP 281
637 ACCCACTTTCAGAGCTGTGGAGATGGCCACAGCACAGGATCTGGATCTG 686
281 heSerIleLeuArgTyrAsnSerThrIleGlnGluSerLeuHisArgSer 297
687 TTGAATACAGAAAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGG 736
298 GluCysProSerGlnArgTyrIleSerLeuGlnCysSerHisCysGlyLe 314
737 CCCTGTCTTCAGGCTCCCTGGTCTCCCTGCATGCTTCCCTGTGGGA 786
314 uArgAlaMetThrGlyArgIleValGlyGlyAlaLeuAlaSerAspSerL 331
787 GAGCCTGAAGACCCCGCTGTGGTGGTGGGAGGAGGCTCTGTGGATT 836
331 ysrPrpProTrpGlnValSerLeuHisPheGlyThrHisIleCysGly 347
837 CTGGGCTTGGCAGGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 886
348 GlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhePh 364
887 GGGAGCATCTGGACCCCACTGGTCTCCTCAGCGCAGCCCACTGCTTC.. 934
364 eValThrArgGluLysValLeuGluGlyTrpLysValThrAlaGlyThrS 381
935AGAAACATACCGATGTGTCACTGGAAGGTGGGGGAGGCTCAG 980
381 erAsnLeuHisGlnLeuProGluAlaAlaSerIleAlaGluIleIle 397

981 ACAAACTGGCAGCTTCCCATCCCTGGCT...GTGGCCAAAGATCATCATC 1027
398AsnSerAsnTyrThrAspGluGluAspAspTyrAspIleAl 411
1028 ATTGAATTCAACCCCATGTAC.....CCCAAGACAAATGACATCGC 1068
411 aLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIleHisProA 428
1069 CCTCATGAAGCTGCAGTCCCACTCACTTTCAGGCACAGTCAGGCCCA 1118
428 laCysLeuProMetHisGlyGlnThrPheSerLeuAsnGluThrCysTrp 444
1119 TCTCTGCTCCCTTCTTCATGAGGAGTCACTCCAGCCACCCCACTCTGG 1168
445 IleThrGlyPheGlyLysThrArgGluThrAspLysLysThrSerProPh 461
1169 ATCATTTGGATGGGCTTTACGAGCAGAAATGGAGGAAGATGTCTGACAT 1218
461 eLeuArgGluValGlnValAsnLeuIleAspPheLysLysCysAsnAspT 478
1219 ACTGCTGCAGGCGTCACTCCAGGTCAATTGACAGCACACGGTGCATGCA 1268
478 yrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAlaGlyAsp 494
1269 ACGATCGCTACCCAGGGGAATCACCAGAGATGATGTGTGCGAGGATC 1318
495 LeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyProLeuVa 511
1319 CCGAAGGGGGTGTGGACACCTGCCAGGGTGACAGTGGTGGCCCTGAT 1368
511 lCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSerTrpGlyT 528
1369 G...TACCAATCTGACAGTGGCATGTGGTGGCATGCTTACGTGGGCT 1415
528 hrGlyCysGlyGlnArgAsnLysProGlyValTyrThrLysValThrGlu 544
1416 ATGCTCGGGGGCCGACCCAGGAGTATACCAAGTCTCAGCC 1465
545 ValLeuProTrpIleTyrSerLysMetGluSerGluVal 557
1466 TATCTCAACTGGATCTACAATGTCTGGAAGGTGAGCTG 1504
seq_name: /cqn2_6/ptodata/2/lna/6B_COMB.seq:US-09-261-416-1
seq_documentation_block:
; Sequence 1, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 144..1511
; OTHER INFORMATION: CDS
US-09-261-416-1
alignment_scores:
Quality: 694.50 Length: 503
Ratio: 2.219 Gaps: 15
Percent Similarity: 62.227 Percent Identity: 33.797
alignment_block:
US-09-879-792-12 x US-09-261-416-1 ..

Align seg 1/1 to: US-09-261-416-1 from: 1 to: 2416

107 ValThrThrSerProThrArgValThrLeuValArgAlaThrProValGI 123
|||||
30 GTACACTGGCCCTACTACTCTCTCCGTGGTGCCATCTACATTTTGGG 79
123 Y.....AlaValProfileArgSerSerProAlaArgSerAlap 136
|
80 ACTCGGAATATTACAGGTAGAGGTGGAGGCGGAGCGGATGTCAGAGGTC 129
136 ro.....AlaThrArgAlaThrArgGluSerProGlyThrSerLeuPro 150
||
130 CTGAATAGTACCATCGGGGAAATATCGCCTGCTGTGAAGCCGCC 179
151 LysPheThrTrpArg..... 155
|||
180 ...TTCTCATCCGATCGCTTTTGGCCCTTGATGATTGAANAATAAGTCC 226
155 155
227 TGTTCACACAGATGCAGATGCTGTGTGCACAGATCCTGTCACTGCTGC 276
156GluGlyGlnLysGlnLeuProLeuLeGlyCysValLeuLeuLeu 170
|||
277 CATTTGAAGTTTTCCTCCCATCATCGTCATTTGGGGATCATTTGCATTGATA 326
171 IleAlaLeuValValSerLeuIleLeuPheGlnPheThrGlnGlyHI 187
:|||||
327 TTACACTGGCCATGTCTGGGCATCCATTCGACGTC..... 365
187 sThrGlyLeuArgTyrLysGluGlnArgGluSerCysProLysHisAlav 204
:|||||
366 .TCAGGAAGTACAGATGTCGCTCATCCTTTAAGTGTACGAGGTGATAA 414
204 aArgCysAspGlyValValAspCysLysLeuLysSerAspGluLeuGly 220
:|||||
415 CTCGATGTACGGAGTCTCGGATTGCAAGACGCGGAGGACGATACCGC 464
221 CysValArgPheAspTrpAspLysSerLeuLeuLysIleTyrSerGlySe 237
:|||||
465 TGTGTCGGGTGGTGTACAGATGCGCGTGTCCAGGTGTTCACAGCTGC 514
237 rSerHisGlnTrpLeuProLysSerSerAsnTrpAsnAspSerTyrS 254
:|||||
515 TTCG.....TGAAGACCATGTGCTCCGATGACTGGAAGGGTCACTACG 558
254 erGluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThr 270
:|||||
559 CAATGTTGCTGTGCCCACTGGGTTCCTCCAGCTATGTGAGTTTCAGAT 608
271 GluValAla.....HisArgAspPheAlaAs 279
:|||||
609 AACCTCAGAGTACGTCGCTGGAGGGCAGTTCCGGGAGGAGTTGTGTC 658
279 nSerPheSerIleLeuArgTyrAsn.....SerThrIleGlnGluSerL 294
:|||||
659 CATCGATCACCTCTTGCAGATACAAAGGTGACTGTCATTACACCACTCAG 708
294 euHis...ArgSerGluCysProSerGlnArgTyrIleSerLeuGlnCys 309
:|||||
709 TATATGAGGAGGGGATGTGCTCTGGCCAGCTGGTTACCTTGCAGTGC 758
310 SerHisCysGly...LeuArgAlaMetThrGlyArgIleValGlyAl 325
:|||||
759 ACAGCCGTGTGCATAGAAGGGGTACAGCTCAGCATCGTGGGTGGA 808
325 aLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPheGlyT 342
:|||||
809 CATGTCTTCTCTCCAGGTGGCCCTGGCAGGCGAGCTTCAGTTCCAGG 858
342 hrThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThr 358
:|||||

859 GCTACACACCTGTGGGGGGCTCTGTCTCATCAGGCCCTGTGGATCATCACT 908
359 AlaAlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLy 375
|||||
909 GCTGCACACTGTGTTTATGACTTG.....TACCTCCCAAGTCAATGAC 952
375 sValTyrAlaGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSerI 392
:|||||
953 CATCCAGGTGGTCTAGTTTCCCTGTTGGCAATCCAGCCCATCCCACT 1002
392 leAlaGlu.....IleIleIleAsnSerAsnTyrThrAspGluGluAsp 407
:|||||
1003 TGGTGGAGAAATGTTTACCACAGCAAGTACAAGCCAAAGAGGCTGGC 1052
408 TyrAspIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHI 424
:|||||
1053 AATGACATCGCCCTTATGAAGCTGGCGGGCCACTCAGTTCAATGAAT 1102
424 sIleHisProAlaCysLeuProMetHisGlyGlnThrPheSerLeuAsnG 441
:|||||
1103 GATCCAGCCTGTGCTGCTGCCAACTCTGAAGAGAACTTCCCGATGGAA 1152
441 luThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLys 457
:|||||
1153 AAGTGTCTGGAGTCAAGATGGGGGCCACAGAGGATGGAGGTGAC... 1199
458 ThrSerProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLy 474
:|||||
1200 GCCTCCCTGTCTGAACACGCGCGCTCTTGTGATTTCCAAACAAGA 1249
474 s.CysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMet 490
:|||||
1250 TCTGCAACACACAGGCGCTGACGGTGGCATCATCTCCCTCCATGCTC 1299
491 CysAlaGlyAspLeuArgGlyGly..ArgAspSerCysGlnGlyAspSer 506
:|||||
1300 TGGCGGGCTACCTGACGGGTGGCGTTGGAACAGCTGCCAGGGGGACAGC 1349
507 GlyGlyProLeuValCysGluGlnAsnArgTrpTyrTyrLeuAlaGlyVa 523
:|||||
1350 GGGGGGGCTGTGTGTCAAGAGAGGAGGCTGTGGAAGTTAGTGGGAGC 1399
523 lThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrT 540
:|||||
1400 GACCAGCTTGGCATCGCTGCCAGACAGCTGAACAGCCTGGGGGTGACA 1449
540 hrLysValThrGluValLeuProTrpIleTyrSerLysMetGluSerGlu 556
:|||||
1450 CCGGTGTCACTCTCTCTGGACTGGATCCACGAGCAGATGGAGAGAGAC 1499
557 ValArg 558
:|||||
1500 CTAATA 1505

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:us-09-510-738A-188

seq_documentation_block:
; Sequence 188, Application US/09510738A
; Patent No. 6268165
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; CURRENT APPLICATION NUMBER: D6223CIP-A
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 188
; SEQ ID NO 188
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
OTHER INFORMATION: full length cDNA of hepsin
US-09-510-738A-188

alignment_scores:
Quality: 642.50 Length: 540
Ratio: 2.040 Gaps: 19
Percent Similarity: 58.333 Percent Identity: 31.852

alignment_block:

US-09-879-792-12 x US-09-510-738A-188 ..

Align seg 1/1 to: US-09-510-738A-188 from: 1 to: 1783

```

54 AlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgAlaSe 70
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1  TCAGCCCGCTTCCAGGACCTACTCTGAGGCCACACAGGTGAGGCA.. 48
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
70 rProGlyArgAlaSerProAlaGlnAlaSerProAlaArgAlaSerProA 87
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
49 .....GCTGGCCTAGCAGGCCCCAC...G 70
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
87 laLeuAlaSerLeuSerArgSerSerSerGlyArgSerSerSerAlaArg 103
   || |||||  |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
71 CCACCCGCTCTGCTCCAGCCGCCGCTGCTGCGGGCCACCATGCTCC 120
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
104 SerAlaSerValThrThrSerProThrArgValThrLeuValArgAlaTh 120
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
121 TGCCCGGCTGGAGCTGACCCGACC.....CCGGCACT 155
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
120 rProValGlyAlaValProIleArgSerSerProAlaArgSerAlaProA 137
   |||  |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
156 ACCTCAGGCTCCGCCCCACCTGCTGGACCCAGGG...TCCACCCCTG 202
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
137 la...ThrArgAlaThrArgGluSer.....ProGlyThrSerLeuProL 151
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
203 GCCCAGGAGGTGACGAGGAATCATTAACAGAGGCAGTGACATGGCGC 252
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
151 ysPheThrTrpArgGluGlyGlnLysGlnLeuPro..... 162
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
253 AG.....AAGGAGGTGGCGGACTGTGCCATGCTGCTCCAGACCC 293
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
163 .....LeuIleGlyCysValLeuLeuLeuLeuAlaLeuVa 174
   |||  |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
294 AAGGTGGCAGCTCTCACTGCGGGGACCTGCTACTTCTGACAGCCATCGG 343
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
174 lValSerLeuIleIleLeuPheGlnPheThrGlnGlyHisThrGlyIleA 191
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
344 GCGGCA.....TCCTGG..... 356
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
191 rgTyrlsGluGlnArgGluSerCysProLysHisAlaValArgCysAsp 207
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
356 ..... 356
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
208 GlyValValAspCysLysLeuLysSerAspGlu.....LeuGlyCysVa 222
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
357 GCCATTGTGCTGTTCTCTCAGAGGTGACACAGGCGCGCTGTACCCAGT 406
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
222 lArgPheAspTrpAspLysSerLeuLeuLysIleTyrlSerGlySerSerH 239
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
407 GCAGGTGAGCTGCGGACCTCGGTCTATGGTCTTTGACAAGACCGAAG 456
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
239 isGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrlSerGlu 255
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
457 GGACGTGGCGGCTGCTGCTGCTCTCGCGCTCCACGCCAGGTAGCCGGA 506
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
256 LysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThrGluVa 272
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
507 CTCAGCTGCGAGAGATGGGCTTCTCAGGGCCTGACCCACTCCGAGCT 556
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
272 lAlaHisArgAsp.....PheAlaAsnS 280
   |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

557 GGAGGTGCGAACGGCGGCCCAATGCACGTGCGGCTTCTTGTGTGG 606
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
280 erPheSerIleLeuArgTyrlAsnSerThrIleGlnGluSerLeuHisArg 296
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
607 ACGAGGGAGGCTGCCCCACACACAGAGGCTGTGGAGTCACTCTCCGTG 656
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
297 SerGluCysProSerGlnArgTyrlSerLeuGlnCysSerHisCysG1 313
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
657 TGTGATTGCCCGCAGAGCGCGTTCTTGGCGCCCATCTGCCAAGACTGTG 706
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
313 yLeuArgAlaMet...ThrGlyArgIleValGlyGlyAlaLeuAlaSerA 329
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
707 CCGCAGGAAGCTGCCCGTGGACCGCATCGTGGAGCGCGGACACCACTG 756
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
329 spSerLysTrpProTrpGlnValSerLeuHisPheGlyThrThrHisIle 345
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
757 TGGCCCGTGGCGTGGCAAGTCAGCTTCGCTATGATGGAGCACACCTC 806
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
346 CysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCy 362
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
807 TGTGGGGATCCCTGCTCTCGGGGACTGGGTGCTGACAGCGGCCACTG 856
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
362 spPhePheValThrArgGluLysValLeuGluGlyTrpLysValTrpAlaG 379
   |||  |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
857 C...TTCCCGAGCGGAACCGGTCTGTCCGATGGCGAGTGTTCGCG 903
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
379 lyThr.....SerAsnLeuHisGlnLeuProGluAlaAlaSerIle 392
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
904 GTCCGTGGCCAGGCTCTCCCCACGGTCTG.....CAGCTGGGGGTG 947
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
393 AlaGluIleIleLeuSerAsnTyr.....ThrAs 403
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
948 CAGGCTGTGCTTACCACGGGGCTATCTTCCCTTTCGGGACCCCAACAG 997
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
403 pGluGluAspAspTyrlAspIleAlaLeuMetArgLeuSerLysProLeuT 420
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
998 CGAGGAGACAGCAACGATATGCGCTGGTCCACCTCTCCAGTCCCCCTGC 1047
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
420 hrLeuSerAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThr 436
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1048 CCCTCACAGAATACATCCAGCCTGTGTGCTCCAGCTGCCGCGCAGGCC 1097
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
437 PheSerLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgG1 453
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1098 CTGGTGGATGGCAAGATCTGTACGTCACGGGCTGGGGCAACACGCACTA 1147
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
453 uThrAspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuI 470
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1148 CTATGGCCAAACAGGCCGG...GTACTCCAGGAGGCTCGAGTCCCCATAA 1194
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
470 leAspPheLysLysCysAsnAspTyrlLeuValTyrlAspSerTyrlLeuThr 486
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1195 TCAGCAATGATGTGCAATGGCGCTGACTTCTATGGAACACAGATCAAG 1244
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
487 ProArgMetMetCysAlaGlyAspLeuArgGlyGlyArgaspSerCysG1 503
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1245 CCCAAGATGTTCTGTGTGCTACCCCGAGGTTGGCATTTGATGCTGCCA 1294
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
503 nGlyAspSerGlyGlyProLeuValCysGlu.....GlnAsnA 516
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1295 GGGCGACGCGGTGGTCCCTTTGTGTGTGAGGACAGCATCTCTCGGAGCG 1344
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
516 snArgTrpTyrlLeuAlaGlyValThrSerTrpGlyThrGlyCysGlyGln 532
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1345 CACGTTGGCGGCTGTGTGGCATTGTGAGTTGGGGCACTGGCTGTGCCCTG 1394
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
533 ArgAsnLysProGlyValTyrlThrLysValThrGluValLeuProTrpI1 549
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1395 GCCCAGAGCAGCGCTCTACACAAAGTCAGTGACTTCCGGGAGTGGAT 1444
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
549 eTyrlSerLysMetGluSer 555
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1445 CTTCCAGGCCATAAAGACT 1463
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-000-846-1

```

seq_documentation_block:
; Sequence 1, Application us/09000846
; Patent No. 5981830
; GENERAL INFORMATION:
; APPLICANT: WU, QINGYU
; APPLICANT: SADLER, JASPER
; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
; DISRUPTED SERINE PROTEASE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,846
; FILING DATE: 30-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/866,058
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: LEBOVITZ, RICHARD M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: BERLX 65P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 35..1282
; US-09-000-846-1

```

```

alignment_scores:
  Quality: 619.50      Length: 380
  Ratio: 2.518         Gaps: 12
  Percent Similarity: 64.737      Percent Identity: 37.105

alignment_block:
US-09-879-792-12 x US-09-000-846-1      ..

Align seg 1/1 to: US-09-000-846-1 from: 1 to: i605

209 ValValAspCysLysLeuLysSerAspGlu.....LeuGlyCysValAr 223
   :::::| | | | | | | | | | | | | | | | | | | | | |
146 ATTGTGACCATCTACTGTCAGAGTGACGAGGCCCACTGTACCAAGTGA 195
   :::::| | | | | | | | | | | | | | | | | | | | | |

223 gPeAsTrpAspLysSerLeuLeuLysLleTySerGlySerSerHisG 240
   :::::| | | | | | | | | | | | | | | | | | | | | |

196 GCTCAGTCCAGGGGACTTCACGGCTTGCGGTGTGTTGACAAAGACGAGGAA 245
   :::::| | | | | | | | | | | | | | | | | | | | | |

```

240 lnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrSerGluLys 256
||| :|||:|||||: |||
246 CGTGAGGCTACTGTGTCCTCCACGCTCCAATGCCAGGGTGCGAGGCCT 295
||| :|||:|||||: |||
257 ThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrThrGluValAla 273
|||:|||||:|||||: |||
296 GCCTGTGAGGAGATGGGCTTTCTCAGGCT CTGGC 330
|||:|||||:|||||: |||
273 ahISArgaspHe AlaSnSerPheSer 282
||| :|||: |||: |||: |||: |||:
331 GCATCGGAGCTGATGTCGCACTGCGGGCGCCAACGCCACATCGGGCT 380
|||:|||||:|||||: |||
283 lleuargtyranSerThrilleGlnGlu 292
|||:|||||:|||||: |||
381 TC TT T T G C G T G G A C G A G G C G G A C T C C G T C T G G C T C A G A G T G C T G G A T 430
|||:|||||:|||||: |||
293 SerLeuHisArgSerGluCysProSerGlnArgTyrlleSerLeuGlnCy 309
|||:|||||:|||||: |||
431 GTCATCTCTGTATGACTGTCCTAGAGGCCGATTCTGTGATGCCACCTG 480
|||:|||||:|||||: |||
309 sSerHiscysGlyLeuArgAlaMet . . . ThrGlyArglleValGlyGlyA 325
|||:|||||: ||| :|||: |||: |||: |||:
481 CCAAGACTGTGGCGCGCAGAAGACTGCGCGTGGACCGCATTCGTGGGGGCC 530
|||:|||||:|||||: |||
325 laLeuAlaSerAspSerLysTrpProTrpGlnValSerLeuHisPhegly 341
|||:|||||:|||||: |||
531 AGGACAGACTCTGGGAAGGTGGCGTGGCAGCTCACGCTCGGTATGAT 580
|||:|||||:|||||: |||
342 ThrThrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuth 358
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
581 GGGACCACCTCTCTGGGGGTCCCTGCTGTCCTGGGACTGGGTCTGCAC 630
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
358 rAlaAlaHiscysPhePheValThrArgGluLysValLeuGluGlyTrpL 375
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
631 TGCTGTCAATTGC . . . TTTCAGACGCGAACCGGTCTCTGTCTCGGTGGC 677
|||:|||||:|||||: |||:|||||:|||||: |||
375 ysValTyralagly ThrSerAsnLeuHisGlnLeuPro 387
|||:|||||:|||||: |||:|||||:|||||: |||
678 GAGTATTGTGCTGCTAGCCCCGACCTCA CCC 712
|||:|||||:|||||: |||:|||||:|||||: |||
388 GluAlaAla SerIleAlaGluIleIleAsnSerAsnTyr . . . 401
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
713 CATCTGTGCAACTGGGGGTTCAGGCTGTGATCATATGGGGCTACT 762
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
402 ThrAspGluGluAspAspTyrAspileAlaLeum 413
||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
763 TCCCTTTCCGAGACCTACTATCGACGAAAACAGCAATGACATTCCTTGG 812
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
413 etArgLeuSerLysProLeuThrLeuSerAlaHisIleHisProAlaCys 429
|||:|||||:|||||: ||| :|||: |||: |||: |||: |||: |||: |||: |||:
813 TCCACCTCTTAGCTCCCTGCTCTCACAGAAATACATCCAGCCAGTGTGT 862
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
430 LeuProMethisGlyGlnThrPheSerLeuAsnGluThrCysTrpIleTh 446
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
863 CTCCTGCTCGGGACAGGCCCTGGTGGATGGCAAGGTCTGTACTGTGAC 912
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
446 rglyPheGlyLysThrArgGluThrAspAspLysThrSerProPheLeuA 463
|||:|||||:|||||: ||| :|||: |||: |||: |||: |||: |||: |||: |||:
913 CGGTGGGGTAACAACA . . . CAGTTCTATGGCCAACAGGCTATGCTGTCTC 959
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
463 rgGluValGlnValAsnLeuIleAspPheLysLysCysAsnAspTyrLeu 479
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
960 ANAGGCCGGGTTCCTCATATACAGAACGAAGTTTGCACAGGCCCGAC 1009
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
480 ValTyrAspSerTyrrLeuThrProArgMetMetCysAlaGlyAspLeuAr 496
||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
1010 TTCTACGGGAATCAGATCAAGCCCCAAGATGTCGTGTGCTACTCTCTGA 1059
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
496 gGlyClyArgAspSerCysGlnGlyAspSerGlyClyProLeuValCysG 513
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
1060 GGGTGGCATTCGCTGCCACAGGGCGACAGTGGAGGCCCTTTCTGTGTGTG 1109
|||:|||||:|||||: |||:|||||:|||||: |||:|||||:|||||: |||
513 luGlnAsn AsnArgTrpTyrrLeuAlaGlyValThrSer 525

LOCATION: 23...2589
OTHER INFORMATION: cDNA sequence of TADG-15
US-09-027-337-1

alignment_scores:
Quality: 584.50 Length: 596
Ratio: 2.022 Gaps: 22
Percent Similarity: 48.490 Percent Identity: 29.195

alignment_block:

US-09-879-792-12 x US-09-027-337-1 ..

Align seg 1/1 to: US-09-027-337-1 from: 1 to: 3147

```

10 SerProAla.....ArgThrPro.....SerAl 17
1188 AGCCCGCGTGGCTGGCGGACCTGCCCAAGGACTACGTGGAGATCAAT 1237
17 aGlyAlaSerProAlaGlnAlaSerProAlaGlyThrProGlyArgA 34
1238 GGGGAGAAATACTGCGGAGAGAGTCCAGTTCGTGTCACCAACAG 1287
34 laSerProAlaGlnAlaSerProAlaGlnAlaSerProAlaGlyThrPro 50
1288 CAACAAGATCACAGTTCGCTTCACCTCAGATCAGTCTACACGACACCG 1337
51 ProGlyArgAlaSerProAlaGlnAlaSerProAla..... 62
1338 GCTTCTAGCTGATACCTCTCTCTAGACTCCAGTGACCCATGCCCGGG 1387
63 .....GlyThrProProGlyArg..... 68
1388 CAGTTCACGTGCGCACGCGGCGGTGTATCCGGAAGGAGCTGCGCTGTA 1437
69 ..AlaSerProGlyArgAlaSerProAlaGlnAlaSerProAlaArgAla 84
1438 TGCGTGGCGGAGTGCACCGACACGACGATGAGCTCAACTGCAGTTGCG 1487
85 SerProAlaLeuAlaSerLeuSerArgSerSerSerGlyArgSerSer 101
1488 ACGCGGCGCCACGAGTTCAGCTGCAAGAACAGTTCGCA...AGCCCTC 1534
101 rAlaArgSerAlaSerVal.....ThrThrSerProThrA 113
1535 TTTCTGGGTGCGACAGTGTGAACGACTGCGGAGACAAACAGCAGCA 1584
113 rgValTyrlLeuValArgAlaThrProValGlyAlaValProIleArgSer 129
1585 GGGGTGCACTGTGCGGCGCCAGACCTTCAGGT...GTTCCAAATGGGAAGT 1631
130 SerProAlaArgSerAlaProAlaThrArgAlaThrArgGluSerProG 146
1632 GCCTCTGAAAGCCAGCAGTGCAGTGGGAGGACACACTGTGGGAGCGG 1681
146 yThrSerLeuProLysPheThrTrpArgGluGlyGlnLysGlnLeuPro 163
1682 TCGGACGAGGCGCTCC.....TGCCCA 1703
163 euIleGlyCysValLeuLeuLeuIleAlaLeuValSerLeuIleIle 179
1704 AGGTGAACCTGCTC..... 1717
180 LeuPheGlnPheTrpGlnGlyHisThrGlyIleArgTyrlLysGluGln 196
1717 ..... 1717
196 gLuserCysProLysHisAlaValArgCys..... 206
1718 ....ACTTGTACAAACACACCTACCGCTGCCTCAATGGCTCTGCTTGA 1763
207 ..AspGlyValValAspCysLysLeuLeuSerAspGluLeuGlyCysVal 222
    ||| :||| |||

```

```

1764 GCAAGGCAACCTCTAGTGTGACGGGAAGGAGGAC..... 1798
223 ArgPheAspTrpAspLysSerLeuLeuLysIleTyrlSerGlySerSerHi 239
1798 ..... 1798
239 sGlnTrpLeuProIleCysSerSerAsnTrpAsnAspSerTyrlSerGlu 256
    ||| :|||
1799 .....TGTAGC.....GACGGCTCAGATGAGA 1820
256 yThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrGluVal 272
    || ||| :|||
1821 AGGACTGCGAC..... 1831
273 AlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrlAsnSerTh 289
1831 ..... 1831
289 rIleGlnGluSerLeuHisArgSerGluCysProSerGlnArgTyrls 306
1831 ..... 1831
306 erLeuGlnCysSerHisCysGlyLeuArgAlaMetThr.....GlyArg 320
    ||| :|||
1832 .....TGTGGCTGCGGTCAATTCACGACAGACGCTCGT 1864
321 ileValGlyGlyAlaLeuAlaSerAspSerLysTrpProTrpGlnValse 337
    ||| :|||
1865 GTTGTGGGGCACGGATGCGGATGAGGGGAGTGGCCCTGGCAGGTAA 1914
337 rLeuHis...PheGlyThrThrHisIleCysGlyGlyThrLeuIleAsp 353
    ||| :|||
1915 CCTGCATGCTCTGGGCAGGCGCACATCTCGGTGCTTCCTCATCTCTC 1964
353 laGlnTrpValLeuThrAlaAlaHisCysPhePheValThrArgGluLys 369
    ||| :|||
1965 CCAACTGGCTGGTCTCTGCCGACACTGTACATCGATGACAGA..... 2008
370 valLeuGluGly.....TrpLysValTyrlAl 378
    ||| :|||
2009 .....GGATTAGGTACTCAGACCCACGACGAGTGGACGCGCTTCCT 2049
378 aGlyThrSerAsnLeuHisGlnLeuProGluAlaAlaSer..... 391
    ||| :|||
2050 GGGC.....TTCACGACAGCAGCCAGCAGCGCCCTGGGTGC 2090
392 .....IleAlaGluIleIleAlaSerAsnTyrlThrAspGlu 404
    ||| :|||
2091 AGGAGCGCAGGCTCAGCGCATCATCTCCACCCCTCTTCAATGACTTC 2140
405 GluAspAspTyrlAspIleAlaLeuMetArgLeuSerLysProLeuThrLe 421
    ||| :|||
2141 ACCTTCGACTATGACATCGCGCTGTGGAGCTGGAGAAACCGCAGAGTA 2190
421 uSerAlaHisIleHisProAlaCysLeuProMetHisGlyGlnThrPheS 438
    ||| :|||
2191 CAGCTCATGCTGGCGCCATCTCGCTGCGGACGCGCTCCCATGCTCTCC 2240
438 erLeuAsnGluThrCysTrpIleThrGlyPheGlyLysThrArgGluThr 454
    ||| :|||
2241 CTCCGCGCAAGGCCATCTGGGTACGCGGTGGGACACACCC...CAGTAT 2287
455 AspAspLysThrSerProPheLeuArgGluValGlnValAsnLeuIleAs 471
    ||| :|||
2288 GAGGCGCAGTGGCGGCTGATCTGCAAAAGGTGAGATCGCGCTCATCAA 2337
471 pPheLysLysCysAsnAspTyrlValTyrlAspSerTyrlThrProA 488
    ||| :|||
2338 CCAGACCACTCGGAGAACCTCTG.....CCGACGACATCAGCGCGC 2381
488 rgMetMetCysAlaGlyAspLeuArgGlyGlyArgaspSerCysGlnGly 504
    ||| :|||
2382 GCATGATGTGCGTGGCTTCTCTCAGCGCGCGGTGGACTCTCTCCAGGGT 2431

```

505 AspSerGlyGlyProLeu...ValCysGluGlnAsnAsnArgTrpTyrLe 520
 |||||
 2432 GATTCGGGGACCCCTGTCACAGGTGGAGCGGATGGCGGATCTTCCA 2481
 520 uAlaGlyValThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProG 537
 |||||
 2482 GCGCGGTGGTGAAGCTGGGAGACGCTGCTGCAGAGAACAGCCAG 2531
 537 lyValTyrThrLysValThrGluValLeuProTrpIle 549
 |||||
 2532 GCGTGTACAAAGCTCCCTCTGTTTCGGGACTGGATC 2569

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-200-900A-1

seq_documentation_block:
 ; Sequence 1, Application US/08200900A
 ; Patent No. 5665566
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
 ; NUMBER OF SEQUENCES: 38
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/200,900A
 ; FILING DATE: 23-FEB-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meinerdt, Maureen C.
 ; REGISTRATION NUMBER: 31,544
 ; REFERENCE/DOCKET NUMBER: GI 5201-FWC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170 X8574
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2581 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-200-900A-1

alignment_scores:
 Quality: 548.50 Length: 376
 Ratio: 2.354 Gaps: 13
 Percent Similarity: 61.968 Percent Identity: 34.309
 alignment_block:
 US-09-879-792-12 x US-08-200-900A-1 ..
 Align seg 1/1 to: US-08-200-900A-1 from: 1 to: 2581
 206 CysAspGlyValValAspCysLysLysLeuLysSerAspGluLeuGlyCysVa 222
 |||||
 1322 TGTGACGGTTCCTCCACTGTAGGATGGCTCAGATGAACACACACTGTGT 1371
 222 laArgPheAspTrpAspLysSerLeuLeuLysIleTyrSerGlySerSerH 239
 |||||
 1372 GCGT.....CTCTCAATGGCAGCAGAC 1394

239 isGln.....TrpLeuProIle 244
 :::
 1395 ACAGCAGCTGGTTGGTGCAGTTCAGGATCCAAAGCATATGGCATGTAGCC 1444
 245 CysSerSerAsnTrpAsnAspSerTyrSerGluLysThrCysGlnGlnLe 261
 ||:::|||||
 1445 TGTGCCAGAACTGGACACCCAGATCTCAGATGATGTGTCTCAGCTGCT 1494
 261 uGlyPheGluSerAlaHisArgThrGluValAlaHisArgAspPheA 278
 ||||:::|
 1495 GGGACTAGGCACTGGAAACTCATCCGTGCCA..... 1525
 278 laAsnSerPheSer.....IleLeuArgTyrAsnSerThr 289
 ::|||
 1526ACCTTTTCTACTGGAGTGGACCATATGTAAATTTAAACACACGA 1570
 290 IleGlnGluSerLeu.....HisArgSerGluCysProSerGlnAr 303
 ::|||
 1571 CCTAATGGCAGCTTAATACCAAGCCCAAGCAGCTGTAGAGGATTC 1620
 303 gTyrIleSerLeuGlnCysSerHis.....CysGlyLeuArgAlaMet 318
 |||||
 1621 ACTGATTTTCTACAATGTAACATAAATCATGTGGGAAAAAACTGCTGA 1670
 318 hr.....GlyArgIleValGlyAlaLeuAlaSerAspSer 330
 ||:::|||||
 1671 CTCAAGAAGTTAGCCCGAAGATTGTCGGAGGAGTACTCCAGAGAAGGA 1720
 331 LysTrpProTrpGlnValSerLeuHisPheGlyThrThrHisIleCysG 347
 |||||
 1721 GCCTGGCCTTGGGTGCTGCTGTATTTGCAGATCAACAGGTCCTCGG 1770
 347 yGlyThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPhe 364
 ||:::|||||
 1771 AGCTTCTCTGTCGAGCAGGATGGCTGGTGTCTGGCGCCCACTGCTGT 1820
 364 heValThrArgGluLysValLeuGluClyTrpLysValTyrAlaGly... 379
 ::|||
 1821 AC...GGGAAAAATATGGAGCCCTCTAAGTGGAAAGCAGTCTAGGCCTG 1867
 380ThrSerAsnLeu.....HisGlnLeuProGluAlaAlaSer 392
 ::|||
 1868 CATATGGCATCAATCTGACTTCTCTCTCAGATA...GAAACTAGTTGAT 1914
 392 eAlaGluIleIleAsnSerAsnTyrThrAspGluGluAspAspTyrA 409
 ||:::|||||
 1915 TGACCAAAATGTCTATAAACCCACACTACAATAAACGGGAGAAACAATG 1964
 409 spIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAlaHisIle 425
 |||||
 1965 ACATTGCCATGATGCATCTTGAATGAAAGTGAACACTACACAGATTATA 2014
 426 HisProAlaCysLeuPromethHisGlyGlnThrPheSerLeuAsnGluTh 442
 ::|||
 2015 CAGCCTATTTGTTTACCAGAGAAATCAAGTTTTCCTCCAGGAAGAT 2064
 442 rCysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLysThrS 459
 ||||
 2065 TTGTTCTATTGCTGGCTGGGG...GCACCTATATATCAAGGTTCTACTG 2111
 459 erPropheLeuArgGluValGlnValAsnLeuIleAspPheLysLysCys 475
 ::|||
 2112 CAGACGTACTGCAAGAAAGTGCAGTCTCCCTTCTATCAATGAGAAATGT 2161
 476 AsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAl 492
 ::|||
 2162 CAACACAGATGCCA...GAATATAACATTACGGAAAAATATGGTGTGTC 2208
 492 aGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyClyP 509
 |||||
 2209 AGGCTATGAACAGCAGGAGGGGTAGATTCTTGTCTCAGGGGGATTCAGGCGGAC 2258
 509 roLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSer 525

1571 COTAATGGCAGCTTAATACTAATCCAGCCAGCCAAAGAGCTGTAGAGGATTC 1620
 303 gtyrileSerLeuGlnCysSerHis.....CysGlyLeuArgAlaMet 318
 1621 ACTGATTTTGCTACATGTAAGTAAGTAACAATCATGTGGGAAAAAACTGGTGA 1670
 318 hr.....GlyArgIleValGlyGlyAlaLeuAlaSerAspSer 330
 1671 CTCAGAAGTTAGCCCGAAGATGTGCGAGGAAGTACTCCAGAGAAGGA 1720
 331 LysTrpProTrpGlnValSerLeuHisPheGlyThrThrHisIleCysGI 347
 1721 GCCTGGCCTGGGTGGTGTCTGCTATTTCAGCATCAACAGAGTCTCGG 1770
 347 yclYThrLeuIleAspAlaGlnTrpValLeuThrAlaAlaHisCysPheP 364
 1771 AGCTTCTGTGTGAGCAGGATTGGCTGGTGTCTGGCCGCCCACTGCGTGT 1820
 364 heValThrArgGluLysValLeuGluGlyTrpLysValTyrAlaGly... 379
 1821 AC...GGGAGAAATATGGAGCGCTCAATGTGGAAGCAGTCTAGGCGCT 1867
 380ThrSerAsnLeu.....HisClnLeuProGluAlaAlaSerII 392
 1868 CATATGGCATCAAAATCTGATCTTCTCCAGATA...GAAACTAGGTGAT 1914
 392 eAlaGluIleIleLeuAsnSerAsnTyrThrAspGluGluAspAspTyrA 409
 1915 TCACCAANTTGCATAAACCCACACTACAATAACGGGAAGAACACAAATG 1964
 409 spIleAlaLeuMetArgLeuSerLysProIleuThrLeuSerAlaHisIle 425
 1965 ACATTGCCATGATGCATCTTGAATGAAAGTGAACACTACACAGATTATATA 2014
 426 HisProAlaCysLeuProMetHisClnLeuThrPheSerLeuAsGluTh 442

2015	CAGCGCTATTTGTTTACCAAGAAAAATCAAGATTTTTCCTCCCGACGAGAAT	2061
442	rCysTrpIleThrGlyPheGlyLysThrArgGluThrAspAspLysThrS	459
	: :	
2065	TGTGTTCTATTCTGCTGGTGGGG...GCACITATATATCAAGGTTCTACTG	2111
459	erProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysCys	475
	: :	
2112	CAGCGCTACTCAAGAAGACTGACGCTTCCTCCCTTCTATCAAAATGAGAATGT	2161

459 erProPheLeuArgGluValGlnValAsnLeuIleAspPheLysCys 475
:: |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
2112 CAGACGTACTCGAAGAAGTGCAGTCCTCCCTTTATCAANTGAAAAATGT 2161

```

476 AsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCysAl 492
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
2162 CAACAACAGATGCCA...GAATATAACATTACGGAAAAATATGGTGTGC 2208

492 aGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyGlyp 509
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2209 AGGCTATGAACAGCAGGAGGGGTAGATTCTTCAGAGGGGATTCAGGCGGAC 2258

509 roLeuValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValThrSer 525
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2259 CACTCATGTGCCAGAAAAACAACAGACTGGCTCTGGCTGGCTGGCTGCACTCA 2308

526 TrpGlyThrGlyCysGlnArgAsnAlaLysProGlyValTyrThrLysVa 542
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
2309 TTTGGATATCAATGTGCATCGCCTAATCGCCAGGGGTGTATGCCCGGGT 2358

542 lThrGluValLeuProTrpIleTyrSer 551
| : : : : : ||||| |||
2359 CCCAAGGTTTCAGACAGTGGATCAAAAGT 2386

_name: /cgn2_6/ptodata/2/Ina/5B_COMB.seq:US-09-027-337-9
documentation_block:
sequence 9, Application US/09027337B
tent No. 5972616
GENERAL INFORMATION:

```

```
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-09-027-337-9
seq_documentation_block:
; Sequence 9, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
```



```

458 rSerProPheLeuArgGluValGlnValAsnLeuIleAspPheLysLysC 475
   ::::::::::::::::::::::::::::::::::::::::::::
2099 CGGGCTGATCTGCAAAAGGGTGAGATCCGCGTCATCAACACCAACCT 2148
475 ysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetMetCys 491
   ::::::::::::::::::::
2149 GCAGAACCTCTG.....CCGACAGATACACGCCGCGCATGATGTC 2192
492 AlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSerGlyG1 508
   ::::::::::::::::::::
2193 GTGGGCTTCTCAGCGCGCGGTGGACTCTGCCAGGGTGATTCGGGGG 2242
508 yProLeu...ValCysGluGlnAsnAsnArgTrpTyrLeuAlaGlyValT 524
   ::::::::::::::::::::
2243 ACCCTGTCCAGCGTGGAGCGGATGGCGGATCTCCAGGCGGGTGG 2292
524 hrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrThr 540
   ::::::::::::::::::::
2293 TGAGCTGGGAGACGC.TGGCTCAGAGGAACAAGCCAGCGGTGTACACA 2341
541 LysValThrGluValLeuProTrpIle 549
   ::::::::::::::::::::
2342 AGCTCCCTCTGTTTCGGGAATGGATC 2368
seq_name: /cgn2_6/ptodata/2/1na/backfiles1.seq:5200340-5
seq_documentation_block:
;Patent No. 5200340
;APPLICANT: FOSTER, DONALD C.;MULVIHILL, EILEEN R.;O'HARA,
;PATRICK J.;PINGEL, KURT;YOSHITAKE, SHINJI
;TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
;ACTIVATORS
;NUMBER OF SEQUENCES: 34
;CURRENT APPLICATION DATA:
;APPLICATION NUMBER: US/07/53,412
;FILING DATE: 22-MAY-1987
;SEQ ID NO:5
;LENGTH: 1724
5200340-5

```

```

alignment_scores:
  Quality: 520.00      Length: 533
  Ratio: 1.933        Gaps: 22
  Percent Similarity: 50.469      Percent Identity: 29.644

alignment_block:
US-09-879-792-12 x 5200340-5
..
Align seg 1/1 to: 5200340-5 from: 1 to: 1724

131 ProAlaArgSerAlaProAlaThrArgAlaThrArgLysSerProGlyTh 147
   ||||| ||||| ::::::::::::::::::::
165 CAGCAACATCATGATGTCGCGCTGCTGCTCAGAGCAACCGGGTG 214
147 rSerLeuProLysPheThrTrpArgGluGlyGlnLysGlnLeuProLeu 164
   ||||| ||||| ::::::::::::::::::::
215 AATATTGCTGGTGCAACAGTGG.....CAGGGCACAGTGCCTCA 255
164 leGlyCysValLeuLeuLeuIleAlaLeuValSerLeuIleLeu 180
   ::||| ||||| |||||
256 GTGCTGTCAAAAGTTGCAGCGAGCCCAAGTG..... 287
181 PheGlnPheTrpGlnGlyHis.....ThrGly.IleArgTyrLysGlu 195
   ||| ::||| ||||| ::::::::::::::
288 .....TTTCAACGGGGCACCTGCCAGCAGCGCCCTGCTACTTCTCAGAT 331
195 InArgLysSerCysProLys...HisAlaValArgCysAspGlyValVal 210
   ::||| ||||| ::::::::::::::
332 TCGTGTGCCAGTGCCTCCGAGGATTTGCTGGGAAGTGC..... 369
211 AspCysLysLeu..... 214

```

```

   |||:::
370 ...TGTGAATAGATACCAGGGCCACGTGCTACGAGGACCAGGGCATCAG 416
215 .....LysSerAspGluLeuGlyCysValArgPheAspT 226
   :::::: ||| |||||
417 CTACAGGGGCACGTGGAGCACAGCGGAGTGGCGCGAGTGCACCAACT 466
226 rPAspLysSerLeuLeu.....LysIleTyrSerGly..... 236
   |||::: |||||
467 GGAACAGCAGCGGTGGCCCAAGACCCCTACAGCGGGCGGAGGCCAGAC 516
236 ..... 236
517 GCCATCAGGCTGGGCTGGGAAACCACTACTGCAGAAACCCAGATCG 566
237 .SerSerHisGlnTrp..... 241
   ::||| |||
567 AGACTCAAGCCCTGTGCTACGTCTTTAAGGGGGGAAGTACAGCTCAG 616
242 .....LeuProIleCysSerSerAsnTrpAsnAspSerTyrSer 254
   ||| |||||:::
617 AGTTCTGCAGCACCCCTGCTGCTCTGAGGGAACAGTACTGCTAC... 663
255 GluLysThrCysGlnGlnLeuGlyPheGluSerAlaHisArgThrG1 271
   ::||| |||||:::
664 .....TTTGGGAATGGGTACGCTACCGTGGCAGC... 693
271 uValAlaHisArgAspPheAlaAsnSerPheSerIleLeuArgTyrAsnS 288
   ||| :::::: ||| |||||
694 .....CACAGCTCACCGAGTGGGTGCTCTCTGCTCCCGTGGAAAT 736
288 erThrIleGln...GluSerLeuHisArgSerGluCysProSerGlnArg 303
   || ||| ::::::
737 CCATGATCTCTGATAGCAAGTTTACACAGCACAGAACCCAGTGCOCAG 786
304 TyrIleSerLeu..... 307
   ::::::|||
787 GCACGTGGCTGGGCAACATAATTACTGCCGAATCCTGATGGGGATGC 836
307 ..... 307
837 CAAGCCCTGTGCCAGTGTCTGAAGAACCCAGGCTGACGTGGGAGTACT 886
308 .....GlnCysSer.....HisCysGlyLeu..... 314
   ::|||::: |||||
887 GTGATGTGCCCTCTCTGCGCGGTACCTTCATTGATTGTGGGAAGCCTCAA 936
315 .....ArgAlaMetThrGlyArgIleValGlyGlyAlaLeuAlaSe 328
   || ||| |||||:::
937 GTGGAGCCGGAAGAAATGCTCTGGAAGGGTTGTAGGGGGTGTGTGGCCCA 986
328 rAspSerLysTrpProTrpGlnValSerLeu.....HisPheGlyThr 343
   ::||| ||||| ||||| |||||
987 CCACATTCCTGGCCCTGGCAAGTCACTCTAGAACAGGTTTGA...A 1033
343 hrHisIleCysGlyGlyThrLeuIleAspAlaGlnTrpValLeuThrAla 359
   ||||| ||||| ||||| |||||
1034 TGCACITCTGTGGAGCACCTTCATATCCACAGAGTGGGTGTGACTGCT 1083
360 AlaHisCysPhePheValThrArgGluLysValLeuGluGlyTrpLysVa 376
   ||||| ||||| ::::::
1084 GCCCACTGC.....TTGGAGAAGTCCCAAGCCTTCATCTACAGGT 1127
376 lTyrAlaGlyThrSer.....AsnLeuHisGlnLeuProGluAlaA 390
   ||||| ||||| ||||| |||||
1128 CATCTGGGTGGCACACCAAGTGAATCTCGAACCCGATGTTTCAGGAAA 1177
390 laSerIleAlaGluIleIleAsnSerAsnTyrThrAspGluGluAsp 406
   ::||| ||||| ::::::
1178 TAGAAGTGTCTAGGCTGTCTTG.....GAGCCACACA 1209
407 AspTyrAspIleAlaLeuMetArgLeuSerLysProLeuThrLeuSerAl 423
   ||||| ||||| ||||| ||||| |||||

```

```
1210 CGAAAGATATTGCTTGCTAAAGTAAGCAGTCTGCCGTCATCACTGA 1259
423 aHisIleHisProAlaCysLeuProMethHisGlyGlnThrPheSerLeuA 440
::: |||||||
1260 CAAAGTAATCCAGCTGTCTGCCATCCCAAAATATGTGGTGGCTGACC 1309
::: |||||||
440 snGluThrCysTrpIleThrGlyPheGlyLysThrArgGluThrAspAsp 456
::: |||||||
1310 GGACCGAATGTTTCATCACTGCTGGTGGGGAGAAACCCCAAGGTACTTTTGA 1359
457 LysThrSerProPheLeuArgGluValGlnValAsnLeuIleAspPheLy 473
::: |||||||
1360 .....GCTGGCCTTCACAGGAAGCCACGCTCCCTGTGATTGAGATAA 1403
473 sLysCysAsnAspTyrLeuValTyrAspSerTyrLeuThrProArgMetM 490
| ||||| ||| :
1404 AGTGTGCAATCGCTATGAGTTTCTGAATGGAAGAGTCCAATCCACCGAAC 1453
490 etCysAlaGlyAspLeuArgGlyGlyArgAspSerCysGlnGlyAspSer 506
:||||||| ||| ||||| |||||||
1454 TCTGTGCTGGGCATTTGGCCGGAGGCACTGACAGTTGCCAGGTGACAGT 1503
507 GlyGlyProLeuValCysGluGlnAsnAsnArgTyrTyrLeuAlaGlyVa 523
||||||| ||||| :
1504 GGAGGTCTCTGCTTTCGAGAGGACAAATACATTTTACAAGGAGT 1553
523 lThrSerTrpGlyThrGlyCysGlyGlnArgAsnLysProGlyValTyrT 540
||||||| ||||| :
1554 CACTTCTTGGGGTCTTGGCTGTGCACGCCCAATAAGCCTGGTGTCTATG 1603
540 hrLysValThrGluValLeuProTrpIleTyrSerLysMetGluSer 555
:||||| :
1604 TTCGTGTTCAAGGTTTGTTACTTGGATTGAGGGAGTGATGAGAAAT 1650
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 19:40:18 ; Search time 1890.94 Seconds
(without alignments)
12476.693 Million cell updates/sec

Title: US-09-879-792-11
Perfect score: 1748
Sequence: 1 ctcagagaccatggagagg.....ggctgtgtgactcgagaaa 1748

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues
Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	28.2	834	10 BE732381	BE732381 601569601
2	285	16.3	285	9 AW845106	AW845106 MRO-CT000
3	239	13.7	252	10 BF757612	BF757612 ILO-CT300
4	239	13.7	254	10 BF757611	BF757611 ILO-CT300
5	239	13.7	255	10 BF757610	BF757610 ILO-CT300
6	185	10.6	190	9 BE073527	BE073527 RC5-BT055
7	154	8.8	591	9 A1820898	A1820898 zt19a12.y
8	154	8.8	600	9 A1820925	A1820925 zu53f08.y
9	151	8.6	402	10 R78581	R78581 y173c10.r1
c 10	139	8.0	286	9 A1909842	A1909842 QV-BT225-
11	134	7.7	751	10 BE280394	BE280394 601158674
12	103	5.9	235	9 AA402094	AA402094 zu53f08.r
13	101	5.8	384	9 AA285124	AA285124 zt23e06.r
c 14	52	3.0	471	10 R78582	R78582 y173c10.sl
15	49	2.8	614	10 BE286322	BE286322 601095164
16	49	2.8	618	12 BH098720	BH098720 RPCI-24-2
17	49	2.8	622	10 BF119003	BF119003 601755367

005

18	38	2.2	621	10 BE285038	BE285038 601098725
19	38	2.2	678	10 BE289529	BE289529 601088060
20	29	1.7	563	10 BE290088	BE290088 601089056
c 21	24	1.4	317	10 BF750399	BF750399 RCI-BN041
22	24	1.4	523	9 AA542994	AA542994 ni55b01.s
23	24	1.4	527	9 AA101043	AA101043 zm27e11.r
24	24	1.4	619	10 W58737	W58737 zt23607.r1
25	24	1.4	711	9 AU134435	AU134435 AU134435
26	24	1.4	914	10 BI601676	BI601676 603249916
27	24	1.4	1169	10 BG677775	BG677775 602625451
28	23	1.3	688	10 BE309103	BE309103 601093231
29	22	1.3	940	10 BE693852	BE693852 603342491
30	21	1.2	130	9 BE075424	BE075424 MK2-BT058
c 31	21	1.2	237	9 AA861735	AA861735 ak38b11.s
c 32	21	1.2	299	9 AA397523	AA397523 zt72c07.s
c 33	21	1.2	336	9 AW326772	AW326772 19750 MAR
c 34	21	1.2	415	10 BE909673	BE909673 601501682
35	21	1.2	473	9 AW659470	AW659470 96665 MAR
c 36	21	1.2	480	9 A1138859	A1138859 qe04d10.x
c 37	21	1.2	482	9 A1138858	A1138858 qe04d09.x
c 38	21	1.2	558	9 AA393577	AA393577 zt72c07.r
c 39	21	1.2	575	12 AZ227042	AZ227042 RPCI-23-6
c 40	21	1.2	655	12 AZ801541	AZ801541 2M0060005
41	21	1.2	752	10 BI827673	BI827673 603074273
c 42	21	1.2	817	10 BF139806	BF139806 601785786
c 43	21	1.2	904	10 BG823531	BG823531 602728910
c 44	21	1.2	906	10 BF691078	BF691078 602247133
c 45	20	1.1	186	9 BB182766	BB182766 BB182766

ALIGNMENTS

RESULT 1
BE732381
LOCUS 601569601F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3844129 5',
DEFINITION mRNA sequence.
ACCESSION BE732381
VERSION BE732381.1 GI:10146373
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM540 row: b column: 02
High quality sequence stop: 817.
Location/Qualifiers
1. 834
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3844129"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University

FEATURES
source

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL0&t2=IL0-CT3002-071100-500-a05&t3=2000-11-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 251.

FEATURES
Source

1..252
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT3002"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 50 a 71 c 74 g 57 t
ORIGIN

Query Match 13.7%; Score 239; DB 10; Length 252;
Best Local Similarity 100.0%; Pred. No. 5.2e-105;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 cctcctcattgcccctggtgttcgctcatcctcttcacagtcgtcccaagcagcgtgttcgctgtgacgg 573
Db 1 CCTCCTCATGTGCGTGTTCGCTCATCTCCCTCTCCAGTCTTGCGAGGCCACAC 60

QY 574 agggatcaagtacaagagcagagggagagcgtgtcccaagcagcgtgttcgctgtgacgg 633
Db 61 AGGGATCAGGTACAGGAGCAGAGGAGAGCTGTCCCAAGCAGCGTGTTCGCTGTGACGG 120

QY 634 ggtgttgactcaagctgaagctgaagcgtgtggctgctgaggtttgactggagcaa 693
Db 121 GGTGTGGAGTCAAGCTGAAGAGTACAGAGTGGCTGCGTGTGAGTTTGACTGGGACAA 180

QY 694 gtctctgtctaaatctactctgggtctctcccatcagtggttcccatctgtagcagca 752
Db 181 GTCTCTGCTTAAATCTACTCTGGGTCTCCCATCAGTGGCTTCCCATCTGTAGCAGCA 239

RESULT 4
BF757611 254 bp mRNA linear EST 12-JAN-2001
LOCUS IL0-CT3002-071100-500-a03 CT3002 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF757611
ACCESSION BF757611.1 GI:12105615
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 254)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL0&t2=IL0-CT3002-071100-500-a03&t3=2000-11-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 253.

FEATURES
source

1..254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT3002"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 51 a 71 c 75 g 57 t
ORIGIN

Query Match 13.7%; Score 239; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 5.3e-105;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 cctcctcattgcccctggtgttcgctcatcctcttcacagtcgtcccaagcagcgtgttcgctgtgacgg 573
Db 1 CCTCCTCATGTGCGTGTTCGCTCATCTCCCTCTCCAGTCTTGCGAGGCCACAC 60

QY 574 agggatcaagtacaagagcagagggagagcgtgtcccaagcagcgtgttcgctgtgacgg 633
Db 61 AGGGATCAGGTACAGGAGCAGAGGAGAGCTGTCCCAAGCAGCGTGTTCGCTGTGACGG 120

QY 634 ggtgttgactcaagctgaagctgaagcgtgtggctgctgaggtttgactggagcaa 693
Db 121 GGTGTGGAGTCAAGCTGAAGAGTACAGAGTGGCTGCGTGTGAGTTTGACTGGGACAA 180

QY 694 gtctctgtctaaatctactctgggtctctcccatcagtggttcccatctgtagcagca 752
Db 181 GTCTCTGCTTAAATCTACTCTGGGTCTCCCATCAGTGGCTTCCCATCTGTAGCAGCA 239

RESULT 5
BF757610 255 bp mRNA linear EST 12-JAN-2001
LOCUS IL0-CT3002-071100-500-a01 CT3002 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF757610
ACCESSION BF757610.1 GI:12105614
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 255)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Other ESTs: zt19a12.x5
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information
This 5' resequenced clone has no previous 5' data to verify this
new read against
Insert Length: 739 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 467.

FEATURES
source

Location/Qualifiers
1. 591
/organism="Homo sapiens"
/db_xref="GDB:5933877"
/db_xref="taxon:9606"
/clone="IMAGE:713566"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 161 a 160 c 126 g 142 t 2 others
ORIGIN

Query Match 8.8%; Score 154; DB 9; Length 591;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 gagggctggaaggtgtacgagcgccaccgacacgttcctgagggagcctcc 1183
|||||
Db 1 GAGGGCTGGAAGGTGTACGGGGCCACCAGCAACCTGCACCGTTCCTGAGGCGCCCTCC 60

QY 1184 attgccgagatcatcaacagcaattacacccgatgaggagcactatgacatcgcc 1243
|||||
Db 61 ATTGCCGAGATCATCAACAGCAATTACACCGATGAGGAGGAGGACTATGACATCGCC 120

QY 1244 ctcatgcggtgtccaaagccctgacctgtccg 1277
|||||
Db 121 CTCATGCGGCTGTCCAAAGCCCTGACCCCTGTCCG 154

RESULT 8
LOCUS AI820925 600 bp mRNA linear EST 09-JUL-1999
DEFINITION zu53f08.y5 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:741735.5, similar to contains TARI.tl TARI repetitive element
;; mRNA sequence.

ACCESSION AI820925
VERSION AI820925.1 GI:5440004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 600)
REFERENCE NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute / National Institute of Dental Research,
TITLE Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT

Other ESTs: zu53f08.s1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: WashU-NCI human EST Project
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 461.

FEATURES
source

Location/Qualifiers
1. 500
/organism="Homo sapiens"
/db_xref="GDB:5941928"
/db_xref="taxon:9606"
/clone="IMAGE:741735"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT 164 a 162 c 128 g 145 t 1 others
ORIGIN

Query Match 8.8%; Score 154; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1124 gagggctggaaggtgtacgagcgccaccgacacgttcctgagggagcctcc 1183
|||||
Db 1 GAGGGCTGGAAGGTGTACGGGGCCACCAGCAACCTGCACCGTTCCTGAGGCGCCCTCC 60

QY 1184 attgccgagatcatcaacagcaattacacccgatgaggagcactatgacatcgcc 1243
|||||
Db 61 ATTGCCGAGATCATCAACAGCAATTACACCGATGAGGAGGAGGACTATGACATCGCC 120

QY 1244 ctcatgcggtgtccaaagccctgacctgtccg 1277
|||||
Db 121 CTCATGCGGCTGTCCAAAGCCCTGACCCCTGTCCG 154

RESULT 9
LOCUS R78581 402 bp mRNA linear EST 09-JUN-1995
DEFINITION Y173c10.r1 Soares placenta NB2HP Homo sapiens cDNA clone
IMAGE:144882.3, similar to SP:HEPS_RAT_Q05511 SERINE PROTEASE
HEPSIN; contains MER22 repetitive element ;; mRNA sequence.

ACCESSION R78581
VERSION R78581.1 GI:854862
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 402)
REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston
R., Williamson, A., Wohldmann, P. and Willson, R.
The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 554

High quality sequence stops: 312

Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Possible reversed clone: polyT not found

Insert Length: 554 Std Error: 0.00

Seq primer: Promega -21ml3

High quality sequence stop: 312.

FEATURES

Location/Qualifiers

source

1..402
/organism="Homo sapiens"
/db_xref="GDB:558601"
/db_xref="taxon:9606"
/clone="IMAGE:144882"
/clone.lib="Soares placenta Nb2Hp"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AACTGGAAGAAATCGCGCGCAGCAATTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT
ORIGIN

76 a 124 c 117 g 82 t 3 others

Query Match

Best Local Similarity 8.6%; Score 151; DB 10; Length 402;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1106 acccgaggagagctgagggctggaaggtgtacgcgggaccagcaacctgcaccag 1165

DB 205 ACCCGGAGAGAGGTCCTGGAGGGCTGGAAGGTGTACGCGGGACCAGCAACCTGCACCA 264

QY 1166 ttgcctgagcgagctccatgccagatcatcatcaacagcaattacacgatgagga 1225

DB 265 TTGCTTGAGCGAGCCTCCATTTGCCGAGATCATCATCAGCAATTACACCGATGAGGAG 324

QY 1226 gacgactatgacatgcctcatcgggctgt 1256

DB 325 GACGACTATGACATGCCTCATCGGCTGT 355

RESULT 10

A1909842/c

LOCUS

QY-BT225-050599-056 BT225 Homo sapiens cDNA, linear EST 30-MAR-2000

DEFINITION

QY-BT225-050599-056 BT225 Homo sapiens cDNA, mRNA sequence.

ACCESSION

A1909842.1 GI:6500522

VERSION

EST.

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 286)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.J.

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL

MEDLINE

COMMENT

sequence tags

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/seg/gethtml.pl?tl=QVat2-QV-BT225-056.html

&t3=050599&t4=1)

Seq primer: puc 18 forward.

Location/Qualifiers

1..286

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone.lib="BT225"

/sex="female"

/dev_stage="Adult"

/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT

51 a 88 c 85 g 62 t

ORIGIN

Query Match

Best Local Similarity 8.0%; Score 139; DB 9; Length 286;

Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1024 gcacttcggcaccacccacatctgtggagcgacgctcattgcccagtcgggtgtctac 1083

DB 219 GCACITCGGCACCCACCCACATCTGTGGAGGCACGCTCAATTGACGCCAGTGGTGCTCAC 160

QY 1084 tgcgcccaactcttcttgtagccgggagaggtctctcggaggcgtggaagtgtacgc 1143

DB 159 TGCCGCCCACTCTTCTTCGTGACCCGGGAGAGGCTCTTGGAGGGCTGGAAGGTGTACGC 100

QY 1144 gggcaccagcaacctgcaccagttgcctgagcgagcctccattgccgagatcatcaaa 1203

DB 99 GGGCACCAACCACTGCACACCTTGCCTGAGGAGGCTCCATTCGCGAGATCATCATCAA 40

QY 1204 cagcaattac 1213

DB 39 CAGCAATTAC 30

RESULT 11

BE280394

LOCUS

601158674F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3505123 5', linear EST 13-JUL-2000

DEFINITION

601158674F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3505123 5', mRNA sequence.

ACCESSION

BE280394

VERSION

BE280394.1 GI:9155397

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 751)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM181 row: d column: 20
High quality sequence start: 6
High quality sequence stop: 513.
Location/Qualifiers

FEATURES

source

1. 751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3505123"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 182 a 221 c 205 g 143 t
ORIGIN

Query Match 7.7%; Score 134; DB 10; Length 751;
Best Local Similarity 100.0%; Pred. No. 7.1e-54;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 299 ggcaggtcatcaccgaggtcgcgcgtgagcaactcccaaccagagtgacctt 358
|||||
DB 171 GCGAGTATATCGCCAGGTACGCTCGGTGACACCTCCCAACAGAGTGACCTT 230
QY 359 gttagagcaacaccagtgggggtgtaaccatcccgatcatctcctgccagggtcagcacca 418
|||||
DB 231 GTTAGAGCAACACCAGTGGGGGCTGTAGCCATCCGATCATCTCTGCCAGGTACACCA 290
QY 419 gcaaccaggggccac 432
|||||
DB 291 GCAACACGAGGGCCAC 304

RESULT 12
AA402094
LOCUS AA402094 235 bp mRNA linear EST 16-MAY-1997
DEFINITION zu53f08.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:741735 5', mRNA sequence.
ACCESSION AA402094
VERSION AA402094.1 GI:2056077
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 235)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-Merck EST project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 201.
Location/Qualifiers

1. 235
/organism="Homo sapiens"
/db_xref="GDB:5941928"
/db_xref="taxon:9606"
/clone="IMAGE:741735"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 48 a 88 c 56 g 43 t
ORIGIN

Query Match 5.9%; Score 103; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 5.9e-39;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1175 gcagctccattgccgagatcatcatcaacagcattacacccgagagacgactat 1234
|||||
DB 49 GCAGCTCATTCGCGAGATCATCATCAACAGCAATACACCGATGAGGAGCAGCTAT 108
QY 1235 gacatgcctccatcgctgctccaaagccctgacccctgctccg 1277
|||||
DB 109 GACATGCGCCCTCATCGCGCTGTCCAAGCCCTGACCCCTGTCCG 151

RESULT 13
AA285124
LOCUS AA285124 384 bp mRNA linear EST 15-MAY-1997
DEFINITION zt23e06.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
IMAGE:713986 5', mRNA sequence.
ACCESSION AA285124
VERSION AA285124.1 GI:1927878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-Merck EST project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham.
Location/Qualifiers

1. 384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:713986"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1559 cgctgtacctggcaggtgtcaccagctggggcaccaggtgtgtggccaga 1607
|||||
Db 369 CGCTGGGTACCTGGCAGGTGTACCAGCTGGGGCACAGGCTGTGCCCAGA 417

Search completed: August 19, 2002, 21:23:38
Job time: 6200 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 16:01:07 ; Search time 1934.86 Seconds
(without alignments)
12193.480 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctcagagaccatggagagg.....ggctgtgtgactcgagaaa 1748

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpi:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	709.4	40.6	834	10 BE732381	BE732381 601569601
2	440.8	25.2	678	10 BE289529	BE289529 601088060
3	411.6	23.5	614	10 BE286322	BE286322 601095164
4	358.8	20.5	622	10 BF119003	BF119003 601755367
5	348	19.9	402	10 R78581	R78581 y173c10.r1
6	335.8	19.2	621	10 BE285038	BE285038 601098725
7	285	16.3	285	9 AW845106	AW845106 MRO-CT000
8	276.4	15.8	286	9 AI909842	AI909842 QV-BT225-
9	266.4	15.2	563	10 BE230088	BE230088 601089056
10	265.4	15.2	751	10 BE280394	BE280394 601158674
11	240.2	13.7	252	10 BF757612	BF757612 IL0-CT300
12	240.2	13.7	254	10 BF757611	BF757611 IL0-CT300
13	240.2	13.7	255	10 BF757610	BF757610 IL0-CT300
14	188.4	10.8	190	9 BE073527	BE073527 RC5-BT055
15	185.2	10.6	688	10 BE309103	BE309103 601093231
16	172.2	9.9	471	10 R78582	R78582 y173c10.s1
17	170.2	9.7	613	10 BM487066	BM487066 pgm2n.pk0

18	169.8	9.7	598	10 BM487712	BM487712 pgm2n.pk0
19	154.8	8.9	591	9 AI820898	AI820898 zt19a12.y
20	154.8	8.9	600	9 AI820925	AI820925 zu53f08.y
21	148.6	8.5	3030	11 AK004939	AK004939 Mus muscu
22	147.4	8.4	975	10 BG288427	BG288427 602388091
23	144.4	8.3	618	12 BH098720	BH098720 RPCI-24-2
24	143.2	8.2	407	9 BE031440	BE031440 130009 MA
25	142.8	8.2	384	9 AA285124	AA285124 zt23e06.r
26	140.8	8.1	235	9 AA402094	AA402094 zu53f08.r
27	140.8	8.1	490	10 BE483122	BE483122 169169 BA
28	138.4	7.9	465	10 BE857738	BE857738 7947f08.x
29	132.4	7.6	529	9 AW058537	AW058537 wx23b08.x
30	129.4	7.4	450	9 AI393270	AI393270 tg09d11.x
31	128.2	7.3	458	10 BI878819	BI878819 fl83d03.x
32	124.6	7.1	417	9 AI673506	AI673506 wf19h04.x
33	119.6	6.8	737	10 BF168078	BF168078 601776336
34	119.4	6.8	648	9 AI924182	AI924182 wn53f11.x
35	119.2	6.8	649	9 AV656299	AV656299 AV656299
36	119.2	6.8	1052	10 BI554641	BI554641 603235988
37	117.8	6.7	659	9 AI924527	AI924527 wn61c07.x
38	116.4	6.7	876	9 AL555870	AL555870 AL555870
39	115.6	6.6	936	9 AL578261	AL578261 AL578261
40	114.6	6.6	401	10 BM089016	BM089016 502654 MA
41	114.2	6.5	689	10 BG966811	BG966811 602834306
42	112.6	6.4	1021	10 BF161225	BF161225 601769474
43	112.2	6.4	646	10 BJ092444	BJ092444 BJ092444
44	112.2	6.4	771	10 BI831151	BI831151 603080610
45	112	6.4	1050	11 AK006271	AK006271 Mus muscu

ALIGNMENTS

RESULT 1

BE732381 834 bp mRNA linear EST 15-SEP-2000
LOCUS 601569601f1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3844129 5',
DEFINITION BE732381 mRNA sequence.
ACCESSION BE732381 GI:10146373
VERSION BE732381.1
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 834)

NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM540 row: b column: 02

High quality sequence stop: 817.

Location/Qualifiers

1. 834

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3844129"

/clone_lib="NIH_MGC_21"

/tissue_type="choriocarcinoma"

/lab_host="DHI0B (phage-resistant)"

/note="Organ (placenta); Vector: pOTB7; Site:1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dt priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAGGAG(G). size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```
BASE COUNT      186 a      246 c      224 g      178 t
ORIGIN

Query Match      40.6%; Score 709.4; DB 10; Length 834;
Best Local Similarity 95.4%; Pred. No. 1.7e-137;
Matches 795; Conservative 0; Mismatches 31; Indels 7; Gaps 6;

QY 664 gctggcgtggtgaggtttgactgggacagtgctctgtcttaaatctactctgggtcttc 723
Db 1 GCTGGGCTGGCTGAGGTTTACTGGGACAAAGTCTCTCTTAAATCTACTCTGGGTCTC 60

QY 724 ccatcagtggtctcccatctgttagcagcaactgaatgactctactctacagagacctg 793
Db 61 CCATCAGTGGCTTCCCATCTGTAGCAGCAACTGGAATGACTCTCTACTCAGAGAAGACCTG 120

QY 784 ccagcagctgggtttcagagtgctccacgggacaaacggaggttgccacacaggaatttgc 843
Db 121 CCAGCAGCTGGGTTTGGAGAGTGTCTCACCGACAACCGAGGTTGCCACACAGGATTTGC 180

QY 844 caacagcttctaacttgagatacaactccacatccaggaagaagcctccacaggtctga 903
Db 181 CAACAGCTTCTCAATCTTGAGATACAACTCCACCATCCAGGAAGGCTCCACAGGCTGA 240

QY 904 atgcccctccacgggttatctccctccagtggttccactcgcgactgagggccatgac 963
Db 241 ATGCCCTTCCACGGGTATATCTCCCTCCAGTGTCCACTGCGGACTGAGGGCCATGAC 300

QY 964 cggcgagatgctggggggcgctggcctcggaatagcaagtggccttggcaagtgtct 1023
Db 301 CGGGCGGATCGTGGGAGGGCGCTGGCTCGGATAGCAAGTGSGCTTGGCAAGTGTCT 360

QY 1024 gcaattcgacacacacatctgtgagggcagctcatgacgccagtgccgtgtcctac 1083
Db 361 GCACCTTCGGCACACCACATCTGTGGAGGCGACGCTCATTTGACGCCACGCTGGGTGCTCAC 420

QY 1084 tgcggccactgtcttcttcgtaccgggaggaaggtccctggagggctgaaggtgacgc 1143
Db 421 TGCGGCCACTGCTTCTGCTGACCCGGGAGAGGTCCTGGAGGGCTGGAAGGTGTACGC 480

QY 1144 gggcaccagcaacctgcacaggttgcctgagggcagcctcattgcgagatcatcatcaa 1203
Db 481 GGGCACCAAGCAA-CTGCACAGTTCCTGAGGCG-CTCCATTGCCGAGATCATCATCAA 538

QY 1204 cagcaattacacagtaggagagacgacta-tgacatgcgctcatgcgctgttccaagc 1262
Db 539 CAGCAATTACACCGATGAGGAGGAGGACGACTATTGACATCGCCCTCATGCGGCTGTTCCAAG 598

QY 1263 cctgaccctgtccgctcacatccacctgcttgctt-ccccatgcagggacagacctt 1321
Db 599 CCCCTGAACCTGTCGTCACATCCACCTCTGCTTGGCTCCCCATGCATGGACAGACCTTT 658

QY 1322 agcctcaatgacacctgtgtgatacagagcttggcaagaccagggagacagatgacaag 1381
Db 659 AGCCTCAATGAGACCTGTTGGATCACAGGCTTTGGCAAGACAGGAGACAGATGAAAG 718

QY 1382 acatcccccttccctcggagaggtgagtgcaatctcatcgacttcaagaataatgaatgac 1441
Db 719 ACATCCCCCTTCTT-CCGGGAGGTGCAGGTCAATCTCATTCGACTTCCAGAAATGCAATGAC 777

QY 1442 tacttggtctatgacagttacttaccacaaaggaatgatgtgtgctgggacct 1494
Db 778 TAAGTGGTCTATGACAGT--ACCTTACCAGAGGATGATGTGTGTGGGGAACCTT 828

RESULT 2
BE289529
LOCUS
DEFINITION 601088060F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3482919 5',
mRNA sequence.
ACCESSION BE289529
```

VERSION
KEYWORDS
SOURCE
ORGANISM

BE289529.1 GI:9170082

house mouse.

Mus musculus

Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus.

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM8513 row: g column: 16

High quality sequence stop: 615.

Location/Qualifiers

1..678

/organism="Mus musculus"

/strain="C57/B6"

/db_xref="taxon:10090"

/clone="IMAGE:3482919"

/clone_lib="NCI_CGAP_Mam5"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH"

BASE COUNT 148 a 207 c 183 g 140 t

ORIGIN

Query Match 25.2%; Score 440.8; DB 10; Length 678;

Best Local Similarity 82.7%; Pred. No. 1.2e-81;

Matches 516; Conservative 0; Mismatches 107; Indels 1; Gaps 1;

QY 320 tcagcctcgtgacaaacctccccaccagagtgtagctttagagcaacaccagtgagg 379

Db 1 TCAGCCTCCACACATCTCTCCCAAGTTCCTCTGGCAGGAGGCCACAGGCGGAG 60

QY 380 gctgtaccatccgcatctctccaggtcagcaccagcaaccagggccacagggag 439

Db 61 GCTGTCCCAATCCGGGATCTCTCTCCAGGTGAGTCTGCTCATCTCTCTTCTTCTTCT 120

QY 440 agccaggtacagcctgcccgaagttcacctggcgaggggccagaagcagctaccgtc 499

Db 121 AGCCAGGTCTCAGTTTCCCAAGTTCCTCTGGCAGGAGGCCACAGGCGGAGTCCACTC 180

QY 500 atcgggtgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 559

Db 181 ATCGGCTGTGTCATCTCTCTCATCAGCCTGGTGTGATCTGCTCATCTCTCTTCTTCTTCT 240

QY 560 tggcagggccacacagggatcaggtacagagcagagagagagagagagagagagagag 619

Db 241 TGGAGAGCCACACAGGATCAAGTACAAAGACACATGGAGGATTTGCCCTATCCACGA 300

QY 620 gttcgtgtgacgggtgtgtgtgactgcaagctgaagagtgacgagctggctgctgag 679

Db 301 GTTCGCTGTGATGGAGTGGTGGAGTGCAGAAATGAAGACGATGAGCTGGGCTGTCTCAGG 360

QY 680 tttgactgggacagctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 739

Db 361 TTCGACTGGGAAAGTCCCT 420

QY 740 atctgtagcagcaactggaatgactcctactcagagaagacctgccagcagctgggttc 799

```

Db 421 CTCTCAGCAGCAGCTGAACGACACTGACTCCAAGAGGACCTGCCACGAGCTGGGATTT 480
Oy 800 gagagttctccacgacacagaggttggccacagggatttggccaaagcttctcaatc 859
Db 481 GACACGGTTACCGAACAACTGAGGTAGCCACACGAGACATCACCAGCAGCTTCTTACTC 540
Oy 860 ttgagatacaactccaccatccaggaagcctccacaggtctgaatgcccctccacagcg 919
Db 541 TCCGAATAACAACACCACCATCCAGGAAGCCTCTACAGGTGCAATGTCCCTT-CCGGCGG 599
Oy 920 tatatctccctccagtggtccac 943
Db 600 TATGTCCTCCCTCCAGTGTCCAC 623

RESULT 3
BE286322 BE286322 614 bp mRNA linear EST 26-OCT-2000
LOCUS 601095164F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3490022 5',
DEFINITION mRNA sequence.
ACCESSION BE286322
VERSION BE286322.1 GI:9164563
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 614)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM8531 row: o column: 15
High quality sequence stop: 569.
Location/Qualifiers
1. 614
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3490022"
/tissue_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 159 a 168 c 161 g 126 t
ORIGIN

Query Match 23.5%; Score 411.6; DB 10; Length 614;
Best Local Similarity 90.1%; Pred. No. 1.4e-75;
Matches 463; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

Oy 1189 cyagatcatatcaacagcaattacaccagtagagagagcactgacatgcacctcat 1248
Db 1 CCAGATCATCATCAGCGCAACTACACAGATGAACAGGATGACTATGACTATGCCTCAT 60
Oy 1249 gcgctgtccagccctgaccctgtccgtccacatccaccctgttgcctcccatgca 1308
Db 61 CAGGCTGTCCAAGCCCTGAGCCCTGTGACATCCACCTGCCTGCCTCCCGATGCA 120

```

```

Oy 1309 tggacagacaccttttagcctcctcaatgagacctgtggatcacaggcttggcgaaccaggga 1368
Db 121 CGGTGAGACCTTTCCGGCCTCAATGAGACCTG-TGGATCACGGC-TTGGCAAAACCAAGGA 178
Oy 1369 gacagatgacaagacatcccccttctccgggaggtgaggtcaatctcatcgacttcaa 1428
Db 179 GACAGATGAGAAGACATCTCCCTTCTCCGAGAGGTTCAGGTCAACCTCATTCACCTCAA 238
Oy 1429 gaaatgcaatgactactgtgtctatgacagtaccttaccaccaagatgatgtgtgtg 1488
Db 239 GAAGTGCATGACTACTTGGTCTATGACAGCTACCTTACCCCAAGGATGATGTGCGCG 298
Oy 1489 ggaccttctgtggggcagagactcctgccaggggagacagcggggggtctctgtgtgta 1548
Db 299 GGATCTACGAGGAGGAGGAGTCTCTGCCAGGAGACAGTGGAGGACCTCTCGTCTGTGA 358
Oy 1549 gcagaaacacctgctggtacctggcaggtgtaccagctggggcacaggtgtggccagag 1608
Db 359 GCAGAACAACTCTGTTACCTGGCAGGTGTCCACAGTGGGACAGGCTGTGGCCAGAA 418
Oy 1609 aaacaaacctgtgtgtacacacaaagtgcacaaagtcttccctggtattacagcaagt 1668
Db 419 AAACAAGCCTGTGTGTACACCAAAAGTGACAGAAGTACTTCCCTGGATTTACAGAAAGAT 478
Oy 1669 ggagagcgaggtgcgatttcagaaaatccctaaacca 1702
Db 479 GGAGAGTGAGGTACGATTCCTCGGAATCTTAAACCA 512

RESULT 4
BE119003 BE119003 622 bp mRNA linear EST 24-OCT-2000
LOCUS 60175367F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3984761 5',
DEFINITION mRNA sequence.
ACCESSION BE119003
VERSION BE119003.1 GI:10958043
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 622)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9187 row: e column: 18
High quality sequence stop: 618.
Location/Qualifiers
1. 622
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3984761"
/tissue_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 135 a 147 c 185 g 155 t

```

1. 402	source	/organism="Homo sapiens"
/db_xref="GDB:558601"		
/db_xref="taxon:9606"		
/clone="IMAGE:144882"		
/clone_lib="Soares placenta Nb2HP"		
/sex="female"		
/dev_stage="placenta obtained at birth (full term)"		
/lab_host="DH10B (ampicillin resistant)"		
/note="Organ: placenta; Vector: pF7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGGAAGAAATTCGCGCGCAGCAATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pF7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. "		
BASE COUNT	76 a 124 c 117 g 82 t	
ORIGIN		
Query Match	19.9%;	Score 348; DB 10; Length 402;
Best Local Similarity	97.8%;	Pred. No. 2.le-62;
Matches 362;	Conservative 0;	Mismatches 7; Indels 1; Gaps 1;
QY 903	aatgccctccacagcggtatatctccctccagtggtccctccactgcgactcagggccatga	962
Db 1	AATGCCCTTCCCAGCGGTATATCTCCCTCCAGTGTTCCTGCGACTGAGGCCCAATGA	60
QY 963	ccgggcgagtcgtgggagggcgctggctcgatagcaagtggccttggcgaagtgcac	1022
Db 61	CCGGCGGATCGTGGGAGGGCGCTCGCTCGGATAGCAAGTGGCTTGGCAAGTGAGCC	120
QY 1023	tgcacttcggaccacccacacatctgtggagcagcctcattgagcccgagtcgagtgctca	1082
Db 121	TGCACATTCGGGACCCACCCACATCTGTGGAGCGACGCTCATTTGACGCCCGAGTGGGTCA	180
QY 1083	ctgcgcgccactgtcttcctcgtg-accgcggagagaggtccctggagggcgctggaagtgtac	1141
Db 181	CTNCCGCCCACTGCTCTCTCTGTGACCCGGAGAGAGTCTCTGGAGGCTGGAAGTGTAC	240
QY 1142	gcgggcaccagcaacctgcaccagttgctgagcgagcagctccattgcgagatcatcatc	1201
Db 241	GGGGGCACCAGCAAGCTGCACCAAGTTGGCTGAGGCAGCGCTCATTTGCCGAGATCATCATC	300
QY 1202	aacacacattacacgagtagagagacacatgatcatcgcctcatgcgctgtccaag	1261
Db 301	ACAGCAATATACCCGATGAGGAGGACGACTATGACATCGCCCTCATGCGGCTGTTCAAG	360
QY 1262	ccctcgacc 1271	
Db 361	NCCCTTGACC 370	
RESULT 6		
BE285038		
LOCUS	601098725F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3497358 5',	EST 26-OCT-2000
DEFINITION	mus musculus.	
ACCESSION	BE285038	
VERSION	BE285038.1	GI:9162286
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	1 (bases 1 to 621)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cqabp-r@mail.nih.gov	

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM851 row: a column: 07
High quality sequence stop: 529.
Location/Qualifiers

FEATURES
source

1. 621
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:3497358"
/clone.lib="NCI CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 146 a 184 c 163 g 128 t

ORIGIN

Query Match 19.2%; Score 335.8; DB 10; Length 621;
Best Local Similarity 80.8%; Pred. No. 8.4e-60;
Matches 489; Conservative 0; Mismatches 107; Indels 9; Gaps 8;

QY 302 aggtcatatccgcaggtccagctcggtcggtgacacacctcccccaaccagagtgatccttgtt 361
Db 1 AGATCATCATCTGCCAGTCCAGCTCCAGCATCTCCCGACAGAGTGATCTGT 60

QY 362 agagcaaacacagtggtgggtgtacccatccgatcatctctgccaggtcagcacagca 421
Db 61 AGAGCAACACCAAGTGGGGCTGTCCCATCCGGGATCTCTGCCAGGTTCAGCACCA 120

QY 422 accagggccaccaggg-agagccaggttacgagctgcccgaattcacctgg-cgggagg 479
Db 121 ACCAGGGCCACAGGGTAGAGCCAGGCTCAGTTTCCCAAGTTCTCCTGGTCAGGAGA 180

QY 480 gccagaagcagctaccgctcatcggtgctgctctctctctctctctctctctctctct 539
Db 181 CCAGAGCAGCTGCCACTCATCGGTGTGTCTATCTCTCATCAGCTGGTGATCTCGC 240

QY 540 tcatactcttccagttcttgg-cagggccacacagggatcaggtacagagagcagagg 598
Db 241 TCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 600

QY 599 agagctgtcccaagcacagcttctcgtgtacgggtgtgtgac-tgcaagctgaagag 657
Db 301 GAGAGTTGCCCTATCCACGAGTTCGCTGTGATGAGTGGTGGACTTGCAAAATGAAGAG 360

QY 658 tgacagctgggctggtgaggtttgactgggacaggtctgtcttaaaatctactctgg 717
Db 361 CGATGAGTGGGCTGTGTGAGTTCGAGTGGGACAGTCCCTCTCTGAAAGTCTACTCTGG 420

QY 718 gtccctc-coatcagtggttcccatctgttagcagcaactggaatgactctactcagaga 776
Db 421 GTCTTCTGGCAGAGTGGTCTCTGTCTGTCAGCAGCAGC-GGAACGACACATGACTCCAAGA 479

QY 777 agacctgccag-agctgggtttcagagtgctaccgagacaacagaggttcccacagg 835
Db 480 GGACCTGCCAGCAAGCTGGGATTTGACAGCGTTTACCAGAACACTGAGGTAGCCACAGA 539

QY 836 gattttgcaacagcttctcaattgttagatacaactccaccatcaggaaagcctccac 895
Db 540 GACATCACCAGCAGCTTCTTAATCTTCGGAACAAA--ACAACATCCAGGAAGGCTCTAC 919

QY 896 aggtc 900

Db 598 AGGTC 602

|||||

RESULT 7

AW845106
LOCUS MRO-CT0006-010699-022 CT0006 Homo sapiens cDNA, mRNA linear EST 19-MAY-2000
DEFINITION MRO-CT0006-010699-022 CT0006 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW845106
VERSION AW845106.1 GI:7940623
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 285)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zaglo,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=et2-MRO-CT0006-010
699-022&t3=1999-06-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 285.
Location/Qualifiers
1. 285
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="CT0006"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 59 a 85 c 90 g 51 t

ORIGIN

Query Match 16.3%; Score 285; DB 9; Length 285;
Best Local Similarity 100.0%; Pred. No. 2.4e-49;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 957 ccataccgggcggatcgtggaggggcgctggcctcgatagcagtgcttgcaag 1016
Db 1 CCATGACGGGGCGGATCTGGGAGGGCGCTGGCTCGGATAGCAAGTGGCTTGGCAAG 60

QY 1017 tgagtctgcattctggcaccaccacatctgtggaggcagctcattgacgccccagtgg 1076
Db 61 TGAGTCTGCATTCGCGACCAACCACATCTGTGGAGGCACGCTCATTTGACGCCAGTGGG 120

QY 1077 tgctcaactcccccactctctctctgtgacccgggagaggtcctcgagggtggaag 1136
Db 121 TGCTCACTGCCGCCCACTCTTCTTCTGTCACCGGAGAGGTCTCTGGAGGGCTTGAAGG 180

Qy	1137	tgtagcggggaccagcaacctgcaccagttgctcgtgaggcagccttcattgccagatca	1196
Db	181	TGTTACGGGGGACCAGCAACCTGCACCAAGTTGCTGTAGGACGCTCCATTGCCAGATCA	240
Qy	1197	tcatacagacaattacacgatgagagagacgactatgacatcg	1241
Db	241	TCATACAGCAATTACACCGATGAGGAGGACGACTATGACATCG	285
RESULT	8		
AI909842/c			
LOCUS	AI909842	286 bp	mRNA linear EST 30-MAR-2000
DEFINITION	QV-BT225-050599-056 BT225 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	AI909842		
VERSION	AI909842.1	GI:6500522	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 286)		
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brustein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/seq/gethtml.pl?tl=QV&t2=QV-BT225-056.html) Seq primer: puc 18 forward. Location/Qualifiers 1..286 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BT225" /sex="female" /dev_stage="Adult" /note="Organ: breast; Vector: puc18; Site:1; SmaI; Site:2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /16 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	51 a	88 c	85 g
ORIGIN			62 t
Query Match	15.8%	Score 276.4;	DB 9; Length 286;
Best Local Similarity	97.9%;	Pred. No. 1.5e-47;	
Matches	280; Conservative	0; Mismatches	6; Indels 0; Gaps 0;
Qy	957	ccatgacggcgagatcgtagggagggcgctgcctcgatagacaagtgccttggcgaag	1016
Db	286	CCATGACCGGGCGAATCGTGAAGGGCGCTGGCTCGAATAGCAAGTGGCTTGGCAAG	227
Qy	1017	tgaagtgcacttcggcaccaccacatctgtggaggcacgcgtcattgacgcccagtg	1076

D	b	226	TGAGTCACGACTTCCGGCACCACCCACATCTGTGTGAGGCACCGCTCATTTGACGCCCACTGGG	167
Q	y	1077	tgtcactgcgcccactgcttcttcctgaccccgaggagtgctggaggtcgggaag	1136
D	b	166	TGCTCACTGCCGCCCACTGCTTCTTCGTGTGACCCGGAGAAGGCTCTTGAGGGCTGGAAG	107
Q	y	1137	tgtacggcgccacagcaacctgcaccagtgtgctgagcagcctccatgccagatca	1196
D	b	106	TGTTACGGCGGGCACAGCAACCTGCACCACTTGTCTGAGGCAGCCTCCATTGCCGAGATCA	47
Q	y	1197	tcatcaacagacaattcacqgatgaggagcagctatgacatcgc	1242
D	b	46	TCATCAACAGCAATTACCCGATGAGGAGGACGACTATGACATCGC	1
R	E	S	U	L
RESULT		9		
BE290088				
LOCUS		BE290088	563 bp	mRNA linear EST 26-OCT-2000
DEFINITION		601089056F1 NCI_CGAP_Mam5 Mus musculus cdna clone IMAGE:3483900 '5'		
		mRNA sequence.		
ACCESSION		BE290088		
VERSION		BE290088.1	GI:9170901	
KEYWORDS		EST.		
SOURCE		house mouse.		
ORGANISM		Mus musculus		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE		1 (bases 1 to 563)		
JOURNAL		NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: eqabps-x@mail.nih.gov Tissue procurement: Lothar Hennighausen Ph.D., Robin Humphreys cdna Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM8515 row: p column: 13 High quality sequence stop: 515. Location/Qualifiers 1 .563 /organism="Mus musculus" /strain="C57/B6" /db_xref="taxon:10090" /clone="IMAGE:3483900" /clone_lib="NCI_CGAP_Mam5" /tissue_type="tumor, gross tissue" /dev_stage="7 months" /lab_host="DH10B" /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NTH"		
B	A	S	C	G
BASE COUNT		113 a	191 c	147 g
ORIGIN				112 t

	Query Match	15.2%	Score 266.4;	DB 10;	Length 563;
	Best Local Similarity	80.1%;	Pred. No. 2.2e-45;		
	Matches 450;	Conservative 0;	Mismatches 96;	Indels 16;	Gaps 11;
Qy	251	ccagcccgggcgcattccggctctgtgcatactttccaggctccatccgcag-gtcaatc	309		
Db	2	CCAGCCCGGGCATCTCCAGC-CAGGCACCACTTCCAGGTCTATCATCGGCAGCATCATC	60		
Qy	310	atccgcaggttcagctcctgggtgacaaactccccaac-cagagtgtacctgttagagcaa	368		
Db	61	ATTCGCCAGGTTCAGCCTTCCACGACATCCTCCCAACSCAGAGTGTACCTTGTGTAGACAA	120		

```
Qy 369 cacca---gtggggctgtaccatccgatcatctctctgaggtcagcagcagcaacca 425
Db 121 CACCAGCTGCGGGTGTCTCCCATTCGGCGATCTCTCCAGGTCAGCAGCCACCA 180
Qy 426 gggccaccagga-gagccaggtacag-cctgcccaggttcacctggcggaggggcca 483
Db 181 GGGCCACAGGGAGAGCCAGGTCAGTCTTCCCAAGTCTCTCTGGCAGGAGACCCA 240
Qy 484 gaagcagtaacgtctcatcggtgctgctctctctctctctctctctctctctctct 543
Db 241 GAGGAGCTGCCACTCATCGGCTGTGTCATCTCTCATCAGCCCTGGTGATCTCGTCAT 300
Qy 544 cactcttccagttctgagcggggccacacaggg-atcaggtacaaagagcagagggaga 602
Db 301 CCTTCTCTTCTTCTGAGAGAGGCCACACAGGGCATCAAGTACAAAGAGCCACTGGAGA 360
Qy 603 gctctccaaagcagcgtctgctgtacgg-gtgtgtgagctcaagctgaa-gagtga 660
Db 361 GTTCCCTTATCCAGCAGTTCGCTGTGATGACGCTGGGAGTCCAAATGAACGAGCGA 420
Qy 661 cgagctgg---gctgogtggagttgactgggacaaagtctctctctta-aaatctactct 715
Db 421 TGAGCTGGAGCTGTGTACAGGCTTCGACTGGGACAGTCCCTCTGACCACTCTACTCT 480
Qy 716 gggctctccatcagtggtctccatctgttagcagcaactggaatgactctactcagag 775
Db 481 GGGTCTTCTGGCGAGTGGCTTCTCTGTGTCAGCAGCAGCTGGAACGACACTGACTCCACC 540
Qy 776 aagacctgacagcagctgggtt 797
Db 541 AGGACCTGCCAGCAGCTGGGCT 562

RESULT 10
LOCUS BE280394 751 bp mRNA linear EST 13-JUL-2000
DEFINITION 601158674F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3505123 5',
mRNA sequence.
ACCESSION BE280394
VERSION BE280394.1 GI:9155397
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NTH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM181 row: d column: 20
High quality sequence start: 6
High quality sequence stop: 513.
Location/Qualifiers
1. .751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3505123"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DHI0B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
```

```
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT	182 a	221 c	205 g	143 t
ORIGIN				
Query Match	15.2%	Score 265.4;	DB 10;	Length 751;
Best Local Similarity	73.3%;	Pred. No. 4e-45;		
Matches	486;	Conservative 0;	Mismatches 141;	Indels 36; Gaps 10;

```
Qy 183 aggcattccagctggttacacctccagggcg-gaatctccagggccggcattctccagcc 241
Db 53 ATGATCTCTCCAGCAGAAACACCTTCAGAGTCGAGCATCTCCAGCCCGCATCTCCAGCT 112
Qy 242 caggcatctcagcccggtcctccgctctggcatcacttccagctcctcacc-gg 300
Db 113 GGGACACCTCCAGCGCGGGCATCTCCAGCCAGGATCATCTTCAGAGTCTCTCATCTCGG 172
Qy 301 caggctatctccgcccaggtcagctcgtgtgacaaacctcccccaaccagagtgtaacctgt 360
Db 173 CAGGTCATCATCCGCGCAGGTGAGCTCGGTGACAACCTCCCAACAGAGTGTACTTGT 232
Qy 361 tagagcaacacagtgggggtgttaccatccgatcatctctcggcaggtcagcaaccagc 420
Db 233 TAGAGCAACACAGTGGGGGCTGTACCCATCCGATCATCTCTGTCAGGTTCAGCACCAGC 292
Qy 421 aaccagggccaccagggagagccaggtacagcctgcccaggttcacct--ggcgagg 478
Db 293 AACCAGGCCACAGTGGAGAGCCCGCAGGTACAGCCTGACCAAGTTCACACTGAGCAGGGAG 352
Qy 479 gggcagaagcagctaccgctcctcgtggtgctgct-ctctctcattgcccctgggtgg- 533
Db 353 GGCCAGAAAGCAGCTACCGACTCATCGGAGTGCAGTGTCTCACTCTCTCAATTGCCCTGGATGG 412
Qy 534 -tttcgtcattcattctctctctcagttctggcagggc-----cacacagggatcaggtga 585
Db 413 TTTACGCTCATCATCTCTCTCCAGTTCTTGGCAGGGCACACAGGGATCAAGGTTCACAAGGA 472
Qy 586 caaggagcagaggagagctgtcccaagcacg-ctgttcctgtgtgacggggtggaggact 644
Db 473 GCAAGATGTGTGAGAGCTTGTCCCAAGCAGCAGCTGTGCGCTTGTGCAGCGGGGTATGG 532
Qy 645 gcaagctgaagagtagcagagctgggctgctgaggtttgactggg----- 689
Db 533 GACTTCCAAAGACTGAAGAGGTGACAGCGCTGTCTAGCTGAGGTATTGACTGGGACC 592
Qy 690 -acaagttctgtttaaaatctactctgggt-ctctccatcagtggttcccatctgtag 747
Db 593 AACGTCTCTGCTTTAAATCTTACTCTGGGTCCCTCCAATCAGTGGGATCCCATCTCTGTAG 652
Qy 748 cagcaact-ggaatgactcctcactcagagaagacctgccagcagctgggtttcagagatg 806
Db 653 CAGCACTGGGAATGTACTCTACTACAGAGAAGACTGCCAGGAGTGGGATCAAGAGAGGT 712
Qy 807 ctc 809
Db 713 CCC 715

RESULT 11
LOCUS BF757612 252 bp mRNA linear EST 12-JAN-2001
DEFINITION IL0-CT3002-071100-500-a05 CT3002 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF757612
VERSION BF757612.1 GI:12105616
KEYWORDS human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 252)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
```

TITLE JOURNAL MEDLINE COMMENT	Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL0&t2=IL0-CT3002- 071100-500-a05&t3=2000-11-07&t4=1) Seq primer: puc 18 forward High quality sequence stop: 251. Location/Qualifiers 1. .252 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT3002" /dev_stage="Adult" /note="Organ: colon; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 50 a 71 c 74 g 57 t	
	BASE COUNT ORIGIN Query Match 13.7%; Score 240.2; DB 10; Length 252; Best Local Similarity 98.8%; Pred. No. 4.7e-40; Matches 242; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
	QY 514 cctctcattgcccctgggttttcgcgtcatcctcttccagttctggcaggccacac 573 Db 1 CCTCCTCATTTGCCCTGGTGGTTTCGGCTCATCATCCTCTTCCAGTTCTGGCAGGGCCACAC 60 QY 574 agggatcaggtacaagagcagagggagagctgtcccaagcacgctgttcgctgtgacgg 633 Db 61 AGGGATCAGGTACAAAGGACGAGGGAGAGCTGTCCCAAGCACGCTGTTCCAGTTTGTGACGG 120 QY 634 ggtggtgactgaagctgaagagtagagctgggctgcgtgaggtttgactggggacaa 693 Db 121 GGTGGTGGACTGAAGCTGAAGAGTAGAGCTGGGGCTGGCTGAGGTTTGACTGGGGACAA 180 QY 694 gtctctgtcttaaaatctactctgggtctcccatcagtggtcttcccatctgtgacgacaa 753 Db 181 GTCTCTGCTTAAATCTACTCTGGGTCTCCCATCAGTGGCTTCCCATCTGTAGCAGCAC 240 QY 754 ctgga 758 Db 241 CAGCA 245	
	RESULT 12 BF757611 LOCUS BF757611 254 bp mRNA linear EST 12-JAN-2001 DEFINITION IL0-CT3002-071100-500-a03 CT3002 Homo sapiens cDNA, mRNA sequence. ACCESSION BF757611 VERSION BF757611.1 GI:12105615 KEYWORDS EST. SOURCE human.	
TITLE JOURNAL MEDLINE COMMENT	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 254) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL0&t2=IL0-CT3002- 071100-500-a03&t3=2000-11-07&t4=1) Seq primer: puc 18 forward High quality sequence stop: 253. Location/Qualifiers 1. .254 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT3002" /dev_stage="Adult" /note="Organ: colon; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplifications were performed under low stringency conditions." 51 a 71 c 75 g 57 t	
	BASE COUNT ORIGIN Query Match 13.7%; Score 240.2; DB 10; Length 254; Best Local Similarity 98.8%; Pred. No. 4.7e-40; Matches 242; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
	QY 514 cctctcattgcccctgggttttcgcgtcatcctcttccagttctggcaggccacac 573 Db 1 CCTCCTCATTTGCCCTGGTGGTTTCGGCTCATCATCCTCTTCCAGTTCTGGCAGGGCCACAC 60 QY 574 agggatcaggtacaagagcagagggagagctgtcccaagcacgctgttcgctgtgacgg 633 Db 61 AGGGATCAGGTACAAAGGACGAGGGAGAGCTGTCCCAAGCACGCTGTTCCAGTTTGTGACGG 120 QY 634 ggtggtgactgaagctgaagtagagctgggctgcgtgaggtttgactggggacaa 693 Db 121 GGTGGTGGACTGAAGCTGAAGAGTAGAGCTGGGGCTGGCTGAGGTTTGACTGGGGACAA 180 QY 694 gtctctgtcttaaaatctactctgggtctcccatcagtggtcttcccatctgtgacgacaa 753 Db 181 GTCTCTGCTTAAATCTACTCTGGGTCTCCCATCAGTGGCTTCCCATCTGTAGCAGCAC 240 QY 754 ctgga 758 Db 241 CAGCA 245	
	RESULT 13 BF757610 LOCUS BF757610 255 bp mRNA linear EST 12-JAN-2001	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 19:47:18 ; Search time 3211.55 Seconds
(without alignments)
11390.010 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctcagagaccatggagaggg.....ggctgctgtgactogagaaa 1748

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hgt_hum:*

31: em_hgt_inv:*

32: em_hgt_other:*

33: em_hgtgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1678	96.0	1689	6	AX149579	AX149579 Sequence
2	1663	95.1	2067	6	AX149581	AX149581 Sequence
3	1621	92.7	2393	9	AB048796	AB048796 Homo sapi
4	1253	71.7	1314	6	AX190543	AX190543 Sequence
5	1253	71.7	1314	6	AX190546	AX190546 Sequence
6	1188	68.0	3324	9	AB048797	AB048797 Homo sapi
7	1049	60.0	2192	9	AK027798	AK027798 Homo sapi
8	1019	58.3	1078	6	AX190545	AX190545 Sequence
9	227	13.0	160222	9	AP002962	AP002962 Homo sapi
10	227	13.0	184926	9	AP000757	AP000757 Homo sapi
11	156	8.9	111094	2	AP000683	AP000683 Homo sapi
12	152	8.7	450	6	AX246661	AX246661 Sequence
13	49	2.8	2125	10	BC010843	BC010843 Mus muscu
14	29	1.7	29	6	AX149587	AX149587 Sequence
15	29	1.7	29	6	AX149595	AX149595 Sequence
16	27	1.5	27	6	AX149589	AX149589 Sequence
17	27	1.5	27	6	AX149591	AX149591 Sequence
18	26	1.5	26	6	AX149588	AX149588 Sequence
19	25	1.4	25	6	AX149596	AX149596 Sequence
20	25	1.4	36	6	AX149592	AX149592 Sequence
21	24	1.4	39	6	AX149601	AX149601 Sequence
22	24	1.4	969	9	HUMSERPROT	L33404 Human strat
23	24	1.4	986	6	A42048	A42048 Sequence 1
24	24	1.4	986	6	AR053966	AR053966 Sequence
25	24	1.4	986	6	AR084988	AR084988 Sequence
26	24	1.4	1089	6	AR103963	AR103963 Sequence
27	24	1.4	1089	6	AR103964	AR103964 Sequence
28	24	1.4	9729	9	AF332583	AF332583 Homo sapi
29	24	1.4	9909	9	AF166330	AF166330 Homo sapi
30	24	1.4	109124	2	AC011483	AC011483 Homo sapi
31	24	1.4	217346	2	AC027602	AC027602 Homo sapi
32	24	1.4	230000	9	AF243527	AF243527 Homo sapi
33	23	1.3	23	23	AX190548	AX190548 Sequence
34	23	1.3	164780	2	AC109583	AC109583 Homo sapi
35	23	1.3	179240	2	AC094602	AC094602 Rattus no
36	22	1.3	22	6	AX149594	AX149594 Sequence
37	22	1.3	2009	3	AY070572	AY070572 Drosophil
38	22	1.3	81092	2	AC014998	AC014998 Drosophil
39	22	1.3	160005	3	AC007176	AC007176 Drosophil
40	22	1.3	168909	2	AC017470	AC017470 Drosophil
41	22	1.3	169289	3	AC007145	AC007145 Drosophil
42	22	1.3	265605	3	AE003662	AE003662 Drosophil
43	22	1.3	299620	3	AE003429	AE003429 Drosophil
44	21	1.2	21	6	AX190547	AX190547 Sequence
45	21	1.2	21	6	AX190549	AX190549 Sequence

ALIGNMENTS

RESULT	1					
LOCUS	AX149579	1689 bp	DNA	linear	PAT 08-JUN-2001	
DEFINITION	Sequence 3 from Patent WO0136604.					
ACCESSION	AX149579					
VERSION	AX149579.1	GI:14348013				
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 1689)					
AUTHORS	Madison, E.L. and Ong, E.O.					
TITLE	Nucleic acids encoding endotheliases, endotheliases and uses thereof					
JOURNAL	Patent: WO 0136604-A 3 25-MAY-2001;					
FEATURES	CORVAS INTERNATIONAL, INC. (US)					
source	Location/Qualifiers					
	1..1689					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
CDS	1..1689					
	/note="DNA encoding a transmembrane serine protease"					

Db 1621 AAGTGCAGAAAGTTCTTCCTGGATTACAGCAAGATGAGA 1663
|||||
RESULT 3
AB048796 2393 bp mRNA linear PRI 22-MAR-2001
LOCUS Homo sapiens mspl mRNA for membrane-type mosaic serine protease,
DEFINITION complete cds.
AB048796
ACCESSION AB048796
VERSION GI:13429969
KEYWORDS
SOURCE Homo sapiens lung cDNA to mRNA, clone_lib:Marathon Ready cDNA
(Clontech).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kim,D.R., Sharmin,S., Inoue,M. and Kido,H.
TITLE Cloning and expression of novel mosaic serine proteases with and
without a transmembrane domain from human lung(1)
JOURNAL Biochim. Biophys. Acta 1518 (1-2), 204-209 (2001)
MEDLINE 21167393
REFERENCE 2 (bases 1 to 2393)
AUTHORS Kim,D.R., Inoue,M. and Kido,H.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2000) Hiroshi Kido, The University of Tokushima,
Division of Enzyme Chemistry, Institute for Enzyme Research,
Kuramoto-cho, Tokushima 770, Japan
[E-mail:kido@er.tokushima-u.ac.jp, Tel:81-88-633-7424,
Fax:81-88-633-7425]
FEATURES
Location/Qualifiers
1..2393
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="lung"
 /clone_lib="Marathon Ready cDNA (Clontech)"
1..87
 /gene="mspl"
88..1833
 /gene="mspl"
CDS
 /codon_start=1
 /transl_except=(pos:1573..1575,aa:His)
 /transl_except=(pos:862..864,aa:Arg)
 /transl_except=(pos:979..981,aa:His)
 /product="membrane-type mosaic serine protease"
 /protein_id="BAB39741.1"
 /db_xref="GI:13429970"
 /translation="MERDGHNASPARTPSAGASPAQASPAGTPPPGRASPAQSPAQA
SPAGTPPGRASPAQSPACTPPGRASPAQASPARASPALASLSRSSGRSSA
RSASTVPTRYLVLRATPDVGAVPIRRSPARSAPATRATRESPGLPKFTWREGOKO
LPLICGVLLILLIVSLILPFWGCHGRIRKEQRESCPHAYRCDCGVDDCKLSDE
LCGRFWDKLLIKYISGSQHWLPICSNWNDSYSEXTROLGFESAHRTTEVAHRD
FANSFLHYNTIOESLHRSHCPQSQRVISLOCSHGCRAMTRIGVGALADSKWPW
QVSLHFTHGICGGTLIDAQWLTAAGCFVFVRKVLGWKVYAGTSNLHLPEAASI
AEIILNSNYTDEEDYDALMLRLSKPLTSAHIHPACLPMHGQTSLNETCWITGFGK
TRETDKTSFPLERVQNLDLPKNCNDLYVDLYLTPRMKAGDLHGDRSDCGDSSGR
PLVCQQNRNRYLAGVTSMGTGCGQRNKPFVITYTKVTEVLPFIYSKMEVRSLOQDTAPS
LGTSGGDPGGGAPRV"
3'UTR 1834..2393
polyA_signal 2343..2348
BASE COUNT 553 a 692 c 627 g 521 t
ORIGIN

Query Match 92.7%; Score 1621; DB 9; Length 2393;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ctgagacattggaggagcaccgcgggaatgatcatctccagcaagaaccttcacgc 60
|||||
Db 78 CTCAGAGACCATTGAGAGGACACGCCGGAATGATCTCCAGCAAGAACACCTTCAGC 137
|||||


```
QY 1141 cggggccaccagcaacctgacacagttgctgagcagcctccattgccagatcatcat 1200
Db 1218 CGGGGACACAGCAAACTGACCAAGTTGCCTGAGGAGCCTCCATTGCCGAGATCATCAT 1277
QY 1201 caacagcaattacaccgatgaggagagcactatgacatgcgcctcatgaggctgtccaa 1260
Db 1278 CAACAGCAATATACACCATGAGGAGGACGACTATGACATGCCCTCATGCGGCTGTCCAA 1337
QY 1261 gcccctgacctgtccgctcacatccaccctgcttgcctcccatgcatgacagacctt 1320
Db 1338 GCCCCTGACCCCTGTCCCTCATCATCCACCTGCTTGCTGCCATGATGCAGACACTT 1397
QY 1321 tagcctcaatgagacctgtcggatcacagcgttttggcaagaccagggagacagatgacaa 1380
Db 1398 TAGCCTCAATGAGACCTGCTGGATACAGAGCTTTGGCAAGACCAGGAGGACAGATGACAA 1457
QY 1381 gacatcccttctccgaggtgcaagtcaatctcatcgacttcaagaaatgaatga 1440
Db 1458 GACATCCCTTCTCCCGGAGGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGA 1517
QY 1441 ctactgtctatgacagttaccttaccccaagatgatgtgctgggacctctgtg 1500
Db 1518 CTACTTGGTCTATGACAGTTACCTTACCCCAAGATGATGTGCTGGGACCTTCTGTG 1577
QY 1501 gggcagagactctcgcagggagacagcgggggacctcttctgtgtgagcagaacaacg 1560
Db 1578 GGGCAGAGACTCTCTGCCAGGGAGACAGCGGGGGGCTCTTGTCTGTGAGCAGAAACCG 1637
QY 1561 ctggtacctgagaggtgtaccagctggggcacagcgtgtggccagagaacaacctgg 1620
Db 1638 CTGCTACCTGGCAGGTGTACCACTGCTGGGACAGAGGTGTGGCCAGAGAAACAACCTGG 1697
QY 1621 tgtgtacacaaagtgcacaaagtcttccctggtatttaccagcaagatggag 1672
Db 1698 TGTGTACACCAAGTGACAGAAGTTCTTCCCTGGATTACAGCAAGATGGAG 1749

RESULT 4
AX190543
LOCUS AX190543
DEFINITION Sequence 14 from Patent WO0136645.
ACCESSION AX190543
VERSION AX190543.1 GI:15143854
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Quinn,K.E., Spytek,K.A., Majumder,K., Vernet,C., Burgess,C.,
Fernandes,E., Taupier,R.J., Rastelli,L. and Herrmann,J.L.
Serine/threonine kinase and serine protease polypeptides and
nucleic acids encoding same
Patent: WO 0136645-A 14 25-MAY-2001;
Curagen Corporation (US)
FEATURES
    Location/Qualifiers
        1..1314
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        1..1266
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC50789.1"
            /db_xref="GI:15143855"
            /translation="MESPTSLPKFTWRGQKQLPLIGCVLLIALVLSLILFQFWQ
            GHTIRYKQRESCPKHVRCDVKLKSDELGCVDFDWDKSLKITYSGSHQWLP
            ICSSNWDYSBKTCQQLGFSAHRTFAHDFANSFSLRYNSTIQESLHSECP
            QYISLQCSHCLAMTGRIVGALASDKWPQVSLHFGTTHICGTLIDQWVLT
            AHCFFVTRKVLGKWKVYAGTSLNHLPEAASTAEIINSYTDDEDDYDIALMLRSK
            PLTSLAHIPACLPHGQTFSLNETCWTGFGKTRTDDKTSFPLREVQVNLIDFKK
            NDYLDVSLTFLPMKAGDLRGSDSCQDGGPLVCEQNNRWLAGVTSWGTGGQR
            NKPGYTKVTELPWYISKMESEVFTKS"
BASE COUNT 288 a 384 c 376 g 266 t
```

ORIGIN

```
Query Match 71.7%; Score 1253; DB 6; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 ggaagagcccagggtacgagcctgcccgaagttcaccttgcgggagggccagagcagctacc 495
Db 3 GGAGAGCCCAGGTACGAGCGCTGCCCAAGTTACCTGGCGGGAGGCCAGAGCAGCTACC 62
QY 496 gctcatcggttgogtgcctcctcattgcctctggttgcctctcatcattcctctccca 555
Db 63 GCTCATCGGTGCGTGCCTCCTCTCATTTGCCCTGGTGGTTTCGCTCATCATCTCTTCCA 122
QY 556 gtctgcagggccacacacagggatcaggtacaagagacagagagagagctgtcccaagca 615
Db 123 GTTCTGCGAGGGCCACACAGGGATCAGGTACAAGGACAGAGGAGAGCTGTGCCAAGCA 182
QY 616 cgctgttcgctgtgacgggggtgggtgagctgcaagctgaagagtgacgagctgggctcgt 675
Db 183 CGCTGTTGCTGTGACGGGTGGTGGAGTCAAGCTGAAGAGTACGAGCTGGGCTGGCT 242
QY 676 gaggttgactggacaaagtctctgtcttaaatctactctggttccctccatcagtgct 735
Db 243 GAGGTTTACTGGGACAAAGTCTCTGCTTAAATCTACTCTGGGTCTCTCCATCATCTGCT 302
QY 736 tcccatctgtagcagcaactggaatgactctactctcagaaagacctgccagagctggg 795
Db 303 TCCCATCTGTAGCAGCAACTGGAATGACTCTCTACTAGAGAGACCTGCCAGCAGCTGGG 362
QY 796 ttctgagagtgctcacccggacaaccggaggttggccacagggattttgccacagactctc 855
Db 363 TTTGAGAGTGCTCACCGGACAAACCGAGGTGTGCCACAGGGATTTTGCCAAACAGCTTCTC 422
QY 856 aatttgatagataaactccaccatccaggaaagcctccacaggtctgaatgccttccca 915
Db 423 AATCTTGAGATACAACCTCCACCATCCAGGAAGCCTCCACAGGTCTGAATGCCCTTCCA 482
QY 916 gcggtatattcctccctcagtgctccactcgagctagggcagccatgaccggcgagctcgt 975
Db 483 CGGTTATATCTCCCTCCAGTGTTCCCACTCGGACTAGGGCCATGACCGGGCGGATCGT 542
QY 976 gggagggcgctggcctcgatagcaagtggtgcttggcaagtgagctgacttcacttcggac 1035
Db 543 GGGAGGGCGCTGSCCTCGGATAGCAAGTGGCTTGGCAAGTGAGTCTGCACITTCGGCAC 602
QY 1036 caccacatctgtggagggcacgtcattgacgcccagtggtgctcactgcgcgccactg 1095
Db 603 CACCACCATCTGTGGAGGCACGCTCATTTGACCGCCAGTGGGTGCTCATCTGCGGCCACTG 662
QY 1096 ctctctgtagccgggaggaagctcctggagggctggaaggtgtacgcgggcaccagcaa 1155
Db 663 CTCTCTCGTACCCGGGAGAGGTCCTGGAGGGGTGGAAGGTGTACCGGGGACACCAACAA 722
QY 1156 cctgcaccagttgctgagggcagcctccattgcccagatcatcatcaacagcaattacac 1215
Db 723 CTGTGACCACTTGCCTGAGGCAGCCTCCATTGCGGAGATCATCATCAACGAATTTACAC 782
QY 1216 cgatgaggagagcactatgacatgcctcctcatgctgaggtgtccaaagccctgacctgtc 1275
Db 783 CGATGAGGAGGACGACTATGACATCGGCCCTCATCGGCTGTCCAAGGCCCTGACCCCTGTC 842
QY 1276 cgctcaatccaccctctgtcctcccatgcatggacagaccttttagcctcaatgagac 1335
Db 843 CGCTCACATCACCCCTGCTTGCCTCCCATTCATTCGACAGACCTTTAGCTCAATGAGAC 902
QY 1336 ctctgtagcacagggcttggcagacacagggagacagatgacaagacatccccctcct 1395
Db 903 CTGCTGATCACAGGCTTTGGCAAGACCAGGAGACAGATGACAAAGACATCCCCCTTCCT 962
QY 1396 ccgggaggtgcaggtcaactctcatcgacttcaagaaatgcaatgactacttggctatga 1455
```


BASE COUNT	221 a	296 c	319 g	242 t	
ORIGIN					
Query Match	58.3%; Score 1019; DB 6; Length 1078;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 1019; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	670	ctgcgtgaggtttgactgggacaaagtctctgttaaaatctactctggtccctcccatca	729		
Db	1078	CTGCGTGAGGTTGACTGGGACAAAGTCTCTGCTTAAATCTACTCTGGGTCCTCCCATCA	1019		
QY	730	gtggttcccatctgttagcagcaactgggaatgactctactctcactcagagaagacctgccagca	789		
Db	1018	GTGGCTTCCCATCTGTAGCAGCAACTGGAATGACTCTCTACTCAGAGAAGACCTGCCAGCA	959		
QY	790	gctgggtttcagagtgctccacggacaaccgaggttggccacagagggattttgccaaacag	849		
Db	958	GCTGGGTTTCGAGAGTGCTCACCGGACAACCGAGGTTGGCCACAGGGATTTTGCCCAACAG	899		
QY	850	cttctcaatcttgagatacaactccaccatccaggaaagcctccacaggtctgaatgcc	909		
Db	898	CTTCTCAATCTTGAGATACAACTCCACCATCCAGGAAGCCTCCACAGGTCTGAATGCC	839		
QY	910	ttccacgggtatatctccctcccaagtgttcccaactcggactcggagggccatgaccggcg	969		
Db	838	TTCCACAGCGTATATCTCCCTCCAGTGTCCCACTCGGACTGAGGGCCATGACCGGGC	779		
QY	970	gatcgtggagggcgctgcctcgatagcaagtggccttggaagtgaagtgcactt	1029		
Db	778	GATCGTGGAGGGCGCTGCCTTCGGATAGCAAGTGGCCTTGCAAGTGAGTCTGCAC	719		
QY	1030	cggcacccaccacatctgtggagcgactcattgacgcccagtggtgctcaactccgc	1089		
Db	718	CGGCACCCACATCTGTGGAGGCACGCTCATGTAGCGCCACGAGTGCGTCTACTCCGC	659		
QY	1090	caactgctctctgtgacccgggagaggtcctgagggtggaaggtgtacgcgggac	1149		
Db	658	CCACTGCTTCTTTCGTGACCGCGGAGAAAGTCTGAGGGCTGGAAGTGTACGCGGGCAC	599		
QY	1150	caqcaactgcaccagttacctgagcgacgtccattgcccagatcatcaacagaa	1209		
Db	598	CAGCAACTGACACAGTTCCCTTGAGGCACCCCTCCATTGCCGAGATCATCAACAGCAA	539		
QY	1210	ttacaccgatgaggaggacactatgacatcgccctcatcggtctgctcaagcccttgac	1269		
Db	538	TTACACCGATGAGGAGGACACTATGACATCGCCTCATGCGGCTGTCCAAGCCCTGAC	479		
QY	1270	ccgtgctcgtcaccatccacctgttgctccctcccatgcatggacagaccttttagctcaa	1329		
Db	478	CTGTCCCGTTCATCTCCACCTGCTTGCTCCCTCCCATGTCATGGACAGACCTTTAGCCTCAA	419		
QY	1330	tgagacctgctggatcacaggtcttggaaagaccaggagacagatgacaagacatcccc	1389		
Db	418	TGAGACCTGCTGGATCACAGGCTTTGGCAAGACACAGGAGACAGATGACAAGACATPCCC	359		
QY	1390	cttctccggggaggtgcagggtcaactctcatcgacttcaagaaatgcaatgactacttgt	1449		
Db	358	CTTCTCCGGGAGGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTACTTGT	299		
QY	1450	ctatgacagttaccttaccctcccaagatgatgtgtgctggggacctctgtggggcagaga	1509		
Db	298	CTATGACAGTTTACCTTATCCCAAGGATGATGTGTGCTGGGGACCTTCGTGGGGCAGAGA	239		
QY	1510	ctcctgccagggagacagcgggggcctctgtctgtaagcagaacaaccgctgtacct	1569		
Db	238	CTCCTGCCAGGAGACAGCGGGGGCCCTTCTGCTGTGAGCAGAAACACCGCTGGTACCT	179		
QY	1570	ggcaggtgtcaccagctggggcacaggtgtgtggcagagaacaacacctgggtgttacac	1629		
Db	178	GSCAGGTGTACCAGCTGGGGCACAGGCTGTGGCCAGAGAAACAACTGGTGTGTACAC	119		
QY	1630	caaatgacagaagtctctccctgatttacagcaagatggagagcgaggtgcgattca	1688		
Db	118	CAAAAGTACAGAAAGTTCTCTCCCTGGATTTCACGCAAGATGGAGAGCGAGTGCATCA	60		
RESULT	9				
AP002962/c					
LOCUS					
DEFINITION	Homo sapiens genomic DNA, chromosome 11q clone:RP11-652L8, complete sequence.				
ACCESSION	AP002962				
VERSION	AP002962.2				
KEYWORDS	HTG.				
SOURCE	Homo sapiens DNA, clone:RP11-652L8.				
ORGANISM	Homo sapiens				
REFERENCE					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.				
JOURNAL	Homo sapiens genomic DNA				
AUTHORS	Published Only in Database (2000) In press				
TITLE	2 (bases 1 to 160222)				
JOURNAL	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.				
AUTHORS	Direct Submission				
TITLE	Submitted (28-NOV-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/.				
JOURNAL	Tel:81-45-503-9111, Fax:81-45-503-9170)				
COMMENT	On Sep 17, 2001 this sequence version replaced gi:11526589.				
FEATURES					
Source	1..160222				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="11"				
	/map="11q"				
	/clone="RP11-652L8"				
BASE COUNT	44819 a	37318 c	37151 g	40934 t	
ORIGIN					
Query Match	13.0%; Score 227; DB 9; Length 160222;				
Best Local Similarity	100.0%; Pred. No. 1.6e-121;				
Matches 227; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	31	gaatgcattccagcaagaacaccttcagctggagcatctccagccagggcatctccagc	90		
Db	33836	GAATGCATCTCAGCAAGAACACCTTCAGCTGGAGCATCTCCAGCCAGGCATCTCCAGC	33777		
QY	91	tgggacacctccagccgggcatctccagccagggcatctccagccagggcatctccagc	150		
Db	33776	TGGGACACCTCCAGCCGGGGCATCTCCAGCCAGGCATCTCCAGCCAGGCATCTCCAGC	33717		
QY	151	tgggacacctccgggcccgggcatctccagccagggcatctccagctgggtacacctccag	210		
Db	33716	TGGGACACCTCCGGCGGGCATCTCCAGCCAGGCATCTCCAGCTGGTACACCTCCAGG	33657		
QY	211	ccgggcatctccagccgggcatctccagccagggcatctccagccagggcatctccagcc	257		
Db	33656	CCGGGCATCTCCAGCCGGGGCATCTCCAGCCAGGCATCTCCAGCC	33610		
RESULT	10				
AP000757/c					
LOCUS					
DEFINITION	Homo sapiens genomic DNA, chromosome 11q clone:RP11-728F11, complete sequences.				
ACCESSION	AP000757				
VERSION	AP000757.4				
KEYWORDS	HTG.				
SOURCE	Homo sapiens DNA, clone:RP11-728F11.				
ORGANISM	Homo sapiens				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (sites)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA
Published Only in DataBase (1999) In press
2 (bases 1 to 184926)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Feb 21, 2001 this sequence version replaced gi:11094157.
FEATURES
source
Location/Qualifiers
1..184926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-728F11"
BASE COUNT 47261 a 47596 c 43757 g 46312 t
ORIGIN

Query Match
Best Local Similarity 13.0%; Score 227; DB 9; Length 184926;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 31 gaatgcatctccagcaagaacaccttcagctggagcatctccagccagcattcccaagc 90
|||||
Db 163269 GAATGTCATCTCCAGCAAGAACACCTTCAGCTGGAGCATCTCCAGCCAGCATCTCCAGC 163210
Qy 91 tgggacacctccagcgccgcatctccagccagcattccagccagcattcccaagc 150
|||||
Db 163209 TGGGACACCTCCAGCGCGGCATCTCCAGCCAGCATCTCCAGCCAGCATCTCCAGC 163150
Qy 151 tgggacacctccagcgccgcatctccagccagcattccagctgtacacctccagc 210
|||||
Db 163149 TGGGACACCTCCAGCGCGGCATCTCCAGCCAGCATCTCCAGCTGTACACCTCCAGC 163090
Qy 211 ccgggcatctccagcgccgcatctccagccagcattccagcc 257
|||||
Db 163089 CCGGGCATCTCCAGCGCGGCATCTCCAGCCAGCATCTCCAGCC 163043

RESULT 11
AP000683/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone CMB9-3G3 map 11q23, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION AP000683
VERSION AP000683.2 GI:8118871
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:CMB9-3G3.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 111094)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 111,094 genomic DNA of 11q23
Published Only in DataBase (1999) In press
2 (bases 1 to 111094)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (08-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997557.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-3G3
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 100488 bases at least Q40
Consensus quality: 105330 bases at least Q30
Consensus quality: 107871 bases at least Q20
Insert size: 109694; sum-of-contigs
Quality coverage: 4.61x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
15 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 17437 contig of 17437 bp in length
17338 32179 contig of 14642 bp in length
32280 47896 contig of 15617 bp in length
47997 57580 contig of 9584 bp in length
57681 67202 contig of 9522 bp in length
67303 73887 contig of 6585 bp in length
73988 80352 contig of 6365 bp in length
80453 85964 contig of 5512 bp in length
86065 90832 contig of 4768 bp in length
90933 95398 contig of 4466 bp in length
95499 99811 contig of 4315 bp in length
99912 104226 contig of 4315 bp in length
104327 107335 contig of 3009 bp in length
107436 109226 contig of 1791 bp in length
109327 111094 contig of 1768 bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 17437: contig of 17437 bp in length
* 17438 17537: gap of 100 bp
* 17538 32179: contig of 14642 bp in length
* 32180 32279: gap of 100 bp
* 32280 47896: contig of 15617 bp in length
* 47897 47996: gap of 100 bp
* 47997 57580: contig of 9584 bp in length
* 57581 57680: gap of 100 bp
* 57681 67202: contig of 9522 bp in length
* 67203 67302: gap of 100 bp
* 67303 73887: contig of 6585 bp in length
* 73888 73987: gap of 100 bp
* 73988 80352: contig of 6365 bp in length
* 80353 80452: gap of 100 bp
* 80453 85964: contig of 5512 bp in length
* 85965 86064: gap of 100 bp
* 86065 90832: contig of 4768 bp in length
* 90833 90932: gap of 100 bp
* 90933 95398: contig of 4466 bp in length
* 95399 95498: gap of 100 bp

COMMENT

KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 450)
AUTHORS	Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D., Kassam,A., Lamsan,G., Lamnac,R., Crkvenjakov,R., Dickson,M., Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and Stache-Crain,B.
TITLE	Human genes and gene expression products
JOURNAL	Patent: WO 016753-A 1591 13-SEP-2001;
FEATURES	Chiron Corporation (US) ; Hyseq Inc. (US)
source	Location/Qualifiers
	1..450
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
BASE COUNT	97 a 166 c 113 g 74 t
ORIGIN	
Query Match	8.7%; Score 152; DB 6; Length 450;
Best Local Similarity	100.0%; Pred. No. 4e-77;
Matches 152; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 ctgagagaccatggagaggagacagccagcggaatgcattctccagcaagaacaccttcagc 60
Db	65 CTCAGAGACCATTGAGAGGGAGACAGCCAGCGGAATGCATCTCCAGCAAGAACACCTTCAGC 124
Qy	61 tggagcatctccagccaggcatctccagctgggagaccctccagcgcggtatctccagc 120
Db	125 TGGAGCATCTCCAGCCAGGACATCTCCAGCTGGGACACCTCCAGCGCGGGCATCTCCAGC 184
Qy	121 ccaggcatctccagccaggcgcatctccagctg 152
Db	185 CCAGGCATCTCCAGCCAGGACATCTCCAGCTG 216
RESULT 13	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS	house mouse
SOURCE	
LOCUS	BC010843
DEFINITION	Mus musculus. Similar to mosaic serine protease, clone IMAGE:3490022, mRNA, partial cds.
ACCESSION	BC010843
VERSION	BC010843.1 GI:14789982
KEYWORDS</	

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Theriia; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
AUTHORS	1 (bases 1 to 2125)
TITLE	Strausberg, R.
JOURNAL	Direct Submission
REMARK	Submitted (12-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabbs-r@mail.nih.gov Tissue Procurement: Iohar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) cDNA Sequencing by: Sequencing Group at the Stanford Human Genome

Web site: <http://www.mgc.stanford.edu>
 Contact: (Dickson, Mark) mcdepaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 7 Row: b Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source Location/Qualifiers
1..2125 /organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3490022"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI-CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..531
/codon_start=1
/product="Similar to mosaic serine protease"
/protein_id="AAH10843.1"
/db_xref="GI:14789983"
/translation="PEAASISQIIINGNYTDEQDDYDIALIRLSKPLTISAHIHPACL
PMHGQFGLNETCWITGFKTKETDEKTSPLREVQVNLIDFKKCDYLVYDSYLT
MMCGDLRGDRDSCQDGGPLVCEQNNRWYLAGVTSWGTGCGQKNKPGYTKVTEVL
PWYRKWESEVRERKS"
BASE COUNT 536 a 502 c 541 g 546 t
ORIGIN

Query Match 2.8%; Score 49; DB 10; Length 2125;
Best Local Similarity 100.0%; Pred. No. 1.5e+16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1559 cgtgtgtacctggcaggtgtcaccagctgggacagctgtgcccaga 1607
|||||
Db 391 CGCTGGTACCTGGCAGGTGTCAACAGCTGGGGCACAGGCTGTGGCCAGA 439

RESULT 14
AX149587/c
LOCUS
DEFINITION Sequence 29 bp DNA linear PAT 08-JUN-2001
ACCESSION AX149587
VERSION AX149587.1 GI:14348021
KEYWORDS
SOURCE
ORGANISM
Synthetic construct.
synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 29)
AUTHORS Madison, E.L. and Ong, E.O.
TITLE Nucleic acids encoding endotheliases, endotheliases and uses
thereof
JOURNAL Patent: WO 0136604-A 11 25-MAY-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source Location/Qualifiers
1..29 /organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide Primer"
BASE COUNT 7 a 4 c 15 g 3 t
ORIGIN

Query Match 1.7%; Score 29; DB 6; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1273 gtccgctcacatccaccctgttgcctcc 1301
|||||
Db 29 GTCCGCTCACATCCACCCTGCTTGCCTCC 1

RESULT 15
AX149595/c
LOCUS
DEFINITION Sequence 29 bp DNA linear PAT 08-JUN-2001

DEFINITION Sequence 19 from Patent WO0136604.
ACCESSION AX149595
VERSION AX149595.1 GI:14348028
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 29)
AUTHORS Madison, E.L. and Ong, E.O.
TITLE Nucleic acids encoding endotheliases, endotheliases and uses
thereof
JOURNAL Patent: WO 0136604-A 19 25-MAY-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source Location/Qualifiers
1..29 /organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide primer"
BASE COUNT 7 a 4 c 15 g 3 t
ORIGIN
Query Match 1.7%; Score 29; DB 6; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1273 gtccgctcacatccaccctgttgcctcc 1301
|||||
Db 29 GTCCGCTCACATCCACCCTGCTTGCCTCC 1
Search completed: August 19, 2002, 22:21:21
Job time: 9243 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 16:07:22 ; Search time 83.96 Seconds
(without alignments)
5113.951 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctccagagaccatggagaggg.....ggctgctgtgactcagagaaa 1748

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193.2	11.1	2038	4	US-09-008-271A-18
2	181.2	10.4	2413	4	US-09-518-046-1
3	177.4	10.1	1077	3	US-08-807-151-2
4	177.4	10.1	1077	4	US-09-478-957-2
5	176.4	10.1	1479	4	US-09-342-749-1
6	175.6	10.0	2479	4	US-09-342-749-29
7	158.2	9.1	2416	4	US-09-261-416-1
8	154.2	8.8	2544	4	US-09-518-046-3
9	148.4	8.5	1783	4	US-09-510-738A-188
10	120.4	6.9	1081	2	US-09-016-366A-22
11	120.4	6.9	1081	2	US-08-978-404B-17
12	120.4	6.9	1137	2	US-09-016-366A-18
13	120.4	6.9	1137	2	US-08-978-404B-13
14	118.8	6.8	1128	2	US-09-016-366A-20
15	118.8	6.8	1128	2	US-08-978-404B-15
16	116.6	6.7	970	1	US-08-148-910-3
17	116.6	6.7	970	1	US-08-448-937A-3
18	116.6	6.7	2033	1	US-08-148-910-14
19	116.6	6.7	2033	1	US-08-448-937A-14
20	116.4	6.7	1081	4	US-09-008-271A-15
21	114.8	6.6	771	4	US-09-079-970A-4
22	114.4	6.5	735	4	US-09-079-970A-1
23	112.4	6.4	1154	2	US-09-016-366A-16
24	112.4	6.4	1154	2	US-08-978-404B-11
25	108	6.2	1108	2	US-09-016-366A-14
26	108	6.2	1108	2	US-08-978-404B-20
27	105.4	6.0	1031	2	US-08-978-404B-1

28	103.6	5.9	1605	2	US-09-000-846-1	Sequence 1, Appli
29	102.8	5.9	1103	2	US-09-016-366A-24	Sequence 24, Appli
30	101.2	5.8	1109	4	US-09-088-651-1	Sequence 1, Appli
31	99.2	5.7	329	1	US-08-148-910-2	Sequence 2, Appli
32	99.2	5.7	329	1	US-08-148-910-13	Sequence 13, Appli
33	99.2	5.7	329	1	US-08-448-937A-2	Sequence 2, Appli
34	99.2	5.7	329	1	US-08-448-937A-13	Sequence 13, Appli
35	98	5.6	1097	2	US-08-978-404B-4	Sequence 4, Appli
36	95	5.4	2422	1	US-08-475-845-1	Sequence 1, Appli
37	95	5.4	2422	1	US-08-327-690-1	Sequence 1, Appli
38	95	5.4	2422	2	US-08-660-289-1	Sequence 1, Appli
39	95	5.4	2422	2	US-08-537-807-1	Sequence 1, Appli
40	95	5.4	2422	2	US-08-871-003-1	Sequence 1, Appli
41	95	5.4	2422	3	US-08-464-233-1	Sequence 1, Appli
42	95	5.4	2422	4	US-09-189-607-1	Sequence 1, Appli
43	95	5.4	2422	4	US-09-378-907-1	Sequence 1, Appli
44	95	5.4	2422	5	PCT-US94-05779-1	Sequence 1, Appli
45	94.4	5.4	2830	2	US-09-010-928B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-008-271A-18
; Sequence 18, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-008-271A-18

. NUMBER OF SEO TO NOS. 153

NUMBER OF
; SEO ID NO 1

; LENGTH: 2413

TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

OTHER INFORMATION: entire cDNA sequence of TADG-12 gene

US-09-518-046-1

Db	819	AGGAGGCGCTCTGTGGAAFTCTTGCCCTTGGCAGGTTCAGCATCCAGTACGACAAACGAGCACG	878
Qy	1044	tctgtgaggcaagcgtcattgacgcgccagtgagtgctcactgcgcgccactgctctcttcg	1103
Db	879	TCGTGTGAGGAGGACATCTCTGGACCCGCCACTGGGTCTCTACGGCAGCCACACTGCTTC	934
Qy	1104	tgaccvvggagaaggttccttggagggctggaaaggtgtacgcvgggcaccagcaacctgcacc	1163
Db	935	--AGGAAACATACCAGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAGACAACTGGGCA	992
Qy	1164	agttgcttgaggcagcgtccattgcccagatcatcatcaaacagcaaatcaaccgatgagg	1223
Db	993	GCCTCC--ATCCCTTGGCTGTGGCCAAAGATCATCATTTGAATTCAAACCCCATGTACC	1049
Qy	1224	aggacactatgacatgcgcctcatgcggtgtccaagccctgacctgtcccctcaca	1283
Db	1050	CCAAAGACATGACATCGCCCTCATGAAGCTGCAGTCCACTCACTTCTTCAGGCACAG	1109
Qy	1284	tccacctgttgcctccccatgcatggaagacacctttagocctcaatgagacctgctgga	1343
Db	1110	TCAGCGCCATCTCTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCACTCTGGA	1169
Qy	1344	tcacaggctttggcaagacagaggagacagatgacaagacatcccccttctccggagg	1403
Db	1170	TCATTGTATGGGCTTTAGAGACGAATGAGGAGGAAGATGCTGACATATCTGTGCAGG	1229
Qy	1404	tgagggtcaactctcatcgacttcaagaaatgcaatgactacttggctctatgacagttacc	1463
Db	1230	CGTCAGTCCAGGTCATTGACACACACGCTGCAATGCAGAGATGCGTACCAGGGGAAG	1289
Qy	1464	ttaccccaaggatgatgtgtctg9gg9acctctgt9gg9gagagacctctctgccaggag	1523
Db	1290	TCACCGAAGAGATGATGTGTGCAGGCATCCCCGGAAGGGGGTGTGGACACCTCTGCCAGGGTG	1349
Qy	1524	acagcggggggcctctgtctgtgacgagacaacacgcgtgtacctggcagtgctacca	1583
Db	1350	ACAGTGTGGGCCCCCTGATGT---ACCAATCTGACCAAGTGGCATGTGGTGGGCAATCGTTA	1406
Qy	1584	gctggggcacaggctgtggccagagaaaaacaaacctgggtgtgtacaccaaaatgacagaag	1643
Db	1407	GCTGGGGCTATGCTCGCGGGGCCCGAGCACCCAGGAGTATACACCAAGGCTCTCAGCCT	1466
Qy	1644	ttcttccctggatttacagcaagatggagagcgaggtgcgattcagaataaacttaaacag	1703
Db	1467	ATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTGTAACTGTGCTGCCCTTTGAC	1526

RESULT 2

RESOLUTION
US-09-518-046-1

02 03 210 040 1
: Sequence 1. Application US/09518046

; Patent No. 6294663

; GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

: APPLICANT: Underwood, Lowell J.

TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed

TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof

; FILE REFERENCE: D6192CIP

; CURRENT APPLICATION NUMBER: US/09/518,046

; CURRENT FILING DATE: 2000-03-02

; EARLIER APPLICATION NUMBER: 09/261,416

; EARLIER FILING DATE: 1999-03-03

Query Match	10.4%;	Score 181.2;	DB 4;	Length 2413;
Best Local Similarity	55.6%;	Pred. No. 4.2e-35;		
Matches 414;	Conservative	0;	Mismatches 318;	Indels 12; Gaps 3;

RESULT

RESOLUTION 3
US-08-807-151-2

US-08-807-131-2
: Sequence 2: Application us/08807151

: Patent No. 6043033

FACE NO. 0043055
; GENERAL INFORMATION:

```

; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO01
; CLONE: 556016
; US-08-807-151-2

Query Match 10.1%; Score 177.4; DB 3; Length 1077;
Best Local Similarity 54.5%; Pred. No. 2.6e-34;
Matches 399; Conservative 0; Mismatches 327; Indels 6; Gaps 2;

QY 969 ggatctggaggcgctggcctcgatagcaagtggccttggcaagtgaagtctgcaact 1028
DB 237 GGATCTGGGGCGGAGAGCGCGCTCCCGGGGCGCTGGCGCTGGCAGGTCAGCCTGCACG 296

QY 1029 tcggcaccaccacatctgtggagcagcgtcattgagcccgctggtgctcaactcgg 1088
DB 297 TCAGAACGTCACGCTGTGGAGGCTCCATCATCACCCCGAGTGGATCGTGACGCCG 356

QY 1089 cccactgtcttctgacccggagagaggtccctggaggcgtggaaggtgtaacgcggca 1148
DB 357 CCCACTGCGTGGAAAAACCTCTTAACAATCCATGGCATTTGGACGGCATTTCCGGGGATT 416

QY 1149 ccagcaacctgcaccagttgcttggagcagcagcctccattgcgaga---tcactatcaaca 1205
DB 417 TGAGACAATCTTTTCATGTTCTTATGGAGCGGATACCAAGTAGAAAAAGTATTTCTCATC 476

QY 1206 gcaattacacgtagagagagactatgacatcgccctcattcgctgctcaagccc 1265
DB 477 CAATATGACTCAAGACCAAGAACAAATGACATTTGCGCTGATGAAGCTCAGAGAGCTC 536

QY 1266 tgacctgtccgtccatccacccctgttgcctccctcattgcatggagacaccttagcc 1325
DB 537 TGACTTTCAAGACCTAGTGAACCACTGTGTCTGCGCCCAACCCAGGATGATGCTGCAGC 596

QY 1326 tcaatgagacctgctggtcacaggcttggcaagcagggagacagatgacaagacat 1385
DB 1326 tcaatgagacctgctggtcacaggcttggcaagcagggagacagatgacaagacat 1385

; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO01
; CLONE: 556016
; US-08-807-151-2

Db 597 CAGAACACGCTCTGCTGGATTTCGGGGTGGGGGCCACCGAGGAGA---AAGGGAAGACCT 653
QY 1386 ccccttctcccgagggtgcaggtcaatctcatcgtacttcaagaaatgcaatgactact 1445
Db 654 CAGAAGTCTGAACGCTGCCAAGGTGCTTCTCATTTGACACACAGATGCAACACGAGAT 713
QY 1446 tggctatgacagttaccttaccctcccaagagatgatgtgtgctggggacacctgctggggca 1505
Db 714 ATGTCATGACACACCTGATCATCACACGACCATGATCTGTGCGCGCTTCCTGCAGGGAACG 773
QY 1506 gagactctccaggagagacagcgggggcctctgtgtgtgagcagaaacaacgcctggt 1565
Db 774 TCGATTCTTCCAGGGTGACACAGTGGAGGCTCTGGTCTGCTGCTGCAAAAGCTTACAGACCAGGAGTGT 833
QY 1566 acctggcaggtgtcaccagctggggcacagcgtgtggcagagaaacaacacctggtgt 1625
Db 834 GGCTGATAGGGGATACAAAGCTGGGGTCTGGCTGTGCTGCTGCAAAAGCTTACAGACCAGGAGTGT 893
QY 1626 acaccaagtgcagagaagttcttccctggattttacagcaagatggagagcaggtgcgat 1685
Db 894 ACGGAATGTGATGTTATTCACGGACTGGATTTCGACAAATGAGGCGACAGCGCTAAT 953
QY 1686 tcagaaaaatcct 1697
Db 954 CCACATGGTCTT 965

RESULT 4
US-09-478-957-2
; Sequence 2, Application US/09478957
; Patent No. 6350448
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,957
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,151
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNO01
; CLONE: 556016
; US-09-478-957-2
```


Qy	1089	cccaactgctctctgtgacccggggaaggtcctctggagggtgtgaaaggtgtacgcgggca	1148
Db	884	cccactgctggaaaaactcttaacaatccatgcgcatgtgacggcatgttcggggattt	943
Qy	1149	ccagcaacctgccaagtgtgctgaggcagctccatgccgaga---taatacaaca	1205
Db	944	tgagacaattttcatgtttctatggacgcggataccaagttagaaaaagtattctcatc	1003
Qy	1206	gcaattacacgatgaggaggacgactatgacatgcgcctcatcgcgtgtccaagccc	1265
Db	1004	caaatatgactccaagaccagaacaatgacatgctgatgaagctgcagaagcctc	1063
Qy	1266	tgacctgtgcgtctcaatccacacctgttgcctcccatgcattgacagacctttagcc	1325
Db	1064	tgatttcaacagacctagtgaaccagttgtctgtccaaacccaggatgatgctgaac	1123
Qy	1326	tcaatgagacctgtgattcacagccttttggcaagaccaggagagacagatgacaagacat	1385
Db	1124	cagaaacagctctgtggatttccgggtggggggccacgcggagaga---adgggggaagacct	1180
Qy	1386	cccccttctccggggaggtgcaggttcaatctcatcgcattccaagaaatgcaatgactact	1445
Db	1181	cagaagtgtcgaacgtgccaaagggtcttctcaattggacacacagatgcaacagcat	1240
Qy	1446	tgttctatgacagttaccttaacccaagatgatgtgtctgggggaaccttcgtgggggca	1505
Db	1241	atgtctatgacaacctgatcacaccagccaatgatctgtgtccgcttcctgcaggggaaacg	1300
Qy	1506	gaactcctgcaggagagacagcgggggcctcttctgtgtgacgacaaacaacgcgttgt	1565
Db	1301	tcatcttgcaggggtgacatggaggggcctctggtcacttcgaagaaacaatatcttggt	1360
Qy	1566	acctggcaggtgtccacagctgggacagcgtgtgtggccagagaaacaaacctgtgtgt	1625
Db	1361	ggctgataggggatacaagctgggttctgcgtgtgtccaaagcttacagaccaggatgt	1420
Qy	1626	acaccaaaagtgcagaaagtcttccctggatttaccagacaagatggagcgagg	1679
Db	1421	acgggaatgtgattgtattcacgcgactggatttatcgacaatdaggggcagcg	1474

RESULT 6

```

US-09-342-749-29
; Sequence 29, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRP52 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342.749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-749-29

```

	Query Match	10.08;	Score 175.6;	DB 4;	Length 2479;
	Best Local Similarity	49.18;	Pred. No. 9.9e-34;		
	Matches 602; Conservative	0;	Mismatches 594;	Indels 30;	Gaps 4;
Qy	499	cacgggtggtgcctcctccatgcccgctggtggttcgcctcatcacctctccagtt	558		
Dd	326	cctgggaacctctcgtggagactgcgtggccgcgtggcctaactctggaagtcatggg	385		

QY	559	ctgdcaggccacacagggatcaggctacaaggagcagagggagctgtcccaagcagc	618
DB	386	cagcaagtgtcccaactctgggataagatgcgactcctcaggtactcattcaaccctc	445
QY	619	tgttcgctgtgacggggtggtgactgcgaagctgaagactgaagactgaagactggctcgtgag	678
DB	446	taactgtgtgtagggcgtgtccacatgcgcccgccggggaggagcagaaatcgtgtgttcg	505
QY	679	gtttgactgggacaagtctctgttataatctactcttggctcctccatcagttgcttcc	738
DB	506	cctctacggaccaactcactcctcagatgtactcatctcagaggaagtctggaccac	565
QY	739	catctgtagcagcaactcggaaatgactctactcactcagagaagacctgccagcagctgggttt	798
DB	566	tgtgtccaaagacgactgtgaacgagactcagggcgggcgctcagggaatggcgcta	625
QY	799	cgagagtgctcacggacaaccgaggtgtcccaacaggatttgcacaagcttctcaat	858
DB	626	taagaataattttactctctagccaagaaatagtgatgacagcgatccaccagctttat	685
QY	859	cttgagatacaactccaccatccagg-----aaagctcccaaggtctga	903
DB	686	gaactgaacacaagtgcgcgaatgtcgatatctataaaaaactgtaccacagtatgc	745
QY	904	atgccttccacgcgttatatctcctccagttgtccactcggactcggactgagggccatgac	963
DB	746	ctgttcttcaaaacagtgttctttagcgtgttagcctcgtgggtccaaacttgaactc	805
QY	964	cgggc-----ggatcgtgggaaggcgctggcctcgatagacaagtggccttgaca	1014
DB	806	aagccgcagacgaggaatgtggcggtgagagcgctcccgggcgctggcctcgacca	865
QY	1015	agtgaigtgcacttggcaccaccacatctgtggaggcagctcatctgacgcccagtg	1074
DB	866	ggtcagcctgcagctccagaacgtccagctgtgcggaggctccatcatcaccccgagtg	925
QY	1075	ggtgtcactgcgcgcactgttcttctgaccgcggagaaaggtcctggagggtcgaaa	1134
DB	926	gatcgtgacgcgcgcactgcgtggaaaaacctcttaacaaacctggtcattggacgcg	985
QY	1135	ggtgtacgcgggaccacgaacctgcacagttgctgaagcagcctcattgccagat	1194
DB	986	attgcggggattttgagacaattcttcatgtttctatggcgggatcaccagtcacaaa	1045
QY	1195	catcatca---acagcaattaccagatgaggagacactatgacatcgccctcatgcg	1251
DB	1046	agtgattctcatccaaattatgactccaagaccagaacaaatgacattgcgtgatgaa	1105
QY	1252	gctgtccaaagcccctgacctgtcccgctcacatccacctgtgtcctcccatctgctg	1311
DB	1106	gctgcagaagcctctgacttcaacgacctagtgaacccagtgctgtctgcccacccagg	1165
QY	1312	acagacotttagctcaatgagacctgtgtgatcacaggctttggcagaacacaggagac	1371
DB	1166	catgatctgcagccagaacagctctgtggtattcccggtggggggccaccaggagaga	1224
QY	1372	agatgaacaagacatcccttctccggagtgagtgagtgcaatctcatcgcattcaagaa	1431
DB	1225	--agggaagacctcagaagtcgtgaacgtgccaaagtgcttctcatctgagacacagag	1282
QY	1432	atgcaatgactctgtgctatgacagttacotctaccccaaggaatgattgtgtcgtggga	1491
DB	1283	atgcaacagcagatagtctatgacaaactgatcacaccagccatgactgtgcccgtt	1342
QY	1492	cctctgtgggggcagagactcctgcagggagacagcggggggcctcttctgtctgagca	1551
DB	1343	cctgcaggggaacgtcgattcttgcaggggtgacagtggaggcgctctggtcactcgaa	1402
QY	1552	gaacacccgtgttaactggcaggtgtcacagctggggcacaggtctgtgccagagaaa	1611
DB	1403	caacaaatctgtgtgctgtaggggatacagctcgtgggtctgtgctgtgcgaagctta	1462
QY	1612	caaaccttggtgttacaccaaaatgacaaagtctctccttgatttacacgaagaatgga	1671

```

Db 1463 cagaccaggagtgtaacgggaatgtgatgtattcacggagctgattatctcagaatgaa 1522
      || ||| || ||||| || |||| | ||||| || ||||| || ||||| || |||||
Qy 1672 gagcgaggtgcgattcagaaaaatcct 1697
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1523 ggcacacggcctaataccacatggtctt 1548
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-09-261-416-1
: Sequence 1, Application US/09261416A
: Patent No. 6291663
: GENERAL INFORMATION:
: APPLICANT: O'Brien, Timothy J.
: APPLICANT: Underwood, Lowell J.
: TITLE OF INVENTION: TADC-12: A No. 6291663el Transmembrane Serine Protease
: TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
: FILE REFERENCE: D6192
: CURRENT APPLICATION NUMBER: US/09/261,416A
: CURRENT FILING DATE: 1999-03-03
: NUMBER OF SEQ ID NOS: 14
: SEQ ID NO 1
: LENGTH: 2416
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: LOCATION: 144..1511
: OTHER INFORMATION: CDS
US-09-261-416-1

```

Db	1308	ctacctgacgggtggcgttggaaacactccacagggggacagcggggggcccttggtgtgt	1367
Qy	1547	gagcagacaacacgcgtggtacctgggcagagtgtcacccagctggggcacaggctgtggccag	1606
Db	1368	caagagaggcgctgtggaagttagtgggagcgaccagctttggcatcgctgcgcagac	1427
Qy	1607	agaaacaaacctggtgtgtatcacccaagtgcacagaagttcttccctggatttacagcaag	1666
Db	1428	gtgacaagcctgggtgtacaccctgtcaectctctctggactggtaccagcagcag	1487
Qy	1657	atggagagcgggtgcgattcagaaaa	1693
Db	1488	atggagagacacataaaacctgaaga	1514

```

RESULT 8
US-09-518-046-3
; Sequence 3, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Woodward, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
US-09-518-046-3

```



```
Qy 1025 cacttgcgacc-----accacatctgtgagcgacgctcattgacgcccagtgg 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 AGAGTCGCGGACCGATAGTGGATGCACTTCTGCGGGGCTCCCTCATCCACCCCACTGG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1076 gtgtcactgcgcgcactgttcttctgtagaccgagagaggtccttgagggtggaag 1135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GTGTGACCGGACGCGCACTG---CGTGGACCGGACGTCAGGATCTGCGCGCCTCAGG 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1136 gtgtacgcgggcaccagcaacctgcacagttgtcgtgagcgagcctcattgcgagatc 1195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GTGCAACTGCGGGAGCAGCCTCTACTACGAGCAGCAGTGTGCGGTCAGCAGATC 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1196 atcatcaacagcaattacacagtagagagacgactatgacatcgccctcactgcggtg 1255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 ATCTGTGACCCACACAGTTCTACACCGCCAGATCGGAGACATCGCCCTGCTGAGCTG 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1256 tcaagaccctgacccctgtccgttcacatccaccctgttgcctcccatgcatggagac 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 GAGGAGCGGTGAAGGTCTCCAGCCATCCACACGTCACCCCTGCCCCCTCGCTCAGAG 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1316 acctttagcctcaatgagacgtctgtagcacaggtttggcaagaccagggagacagat 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 ACCTTCCCCCGGGAGTGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1376 gacaagacatccccctt---cctccggagaggtgcaggtcaatctcatcgactcaagaaa 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 CGCCTTCCACCCGCAATTTCTCTGAAGCAGGTGAAGTGTCCCATTAATGGAACACCAT 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1433 tg-----caatgactacttggctctatgacgttaaccttaccctcccaagatgtgtgct 1486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 TGTGACCGCAAAATACACACTTGGCGCTACACGGGAGACGACGTCGCCATCGCTGAG 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1487 ggggacctgtggtgggc-----agagactcctgccaggagagacgctgggggcct 1537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 GACATGCTGTGCGGGAGCAGCCTTCTACTACAGGACGACGTCGCCATCGTCCGTGAC 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1538 ctgtctgtgagcagaacacccctgtgtacctgtgcaggttaccacagctgggacagagc 1597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 CTGCTGTGCAAGGTGAATGTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1598 tgtgttcagagagaacaaacctgtgtgtacacacaaagtgcagaaagtcttccctgatt 1657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 TGTGCCAGCCCAACCGGCTTGCATCTACACCGCTGCTACCTACTACTTGGACTGGATC 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1658 tacaagaatgagagcgaggtgcattcag 1689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 CACCACATGTCCCAAAAGCGGTGAGTCAG 810
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11

```
US-08-978-404B-17
; Sequence 17, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: FAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
```

```
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1081 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-17
```

```
Query Match 6.9%; Score 120.4; DB 2; Length 1081;
Best Local Similarity 52.7%; Pred. No. 2.1e-20;
Matches 396; Conservative 0; Mismatches 326; Indels 30; Gaps 5;
```

```
Qy 965 gggcgagatcgtggagggcgctgcctcgatagcaagtggccttgagcaagtgtgctg 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 GTGGCATCTGTGGGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1025 cacttgcgacc-----accacatctgtgagcgacgctcattgacgcccagtgg 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 AGAGTCGCGGACCGATAGTGGATGCACTTCTGCGGGGCTCCCTCATCCACCCCACTGG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1076 gtgtcactgcgcgcactgttcttctgtagaccgagagaggtccttgagggtggaag 1135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GTGTGACCGGACGCGCACTG---CGTGGACCGGACGTCAGGATCTGCGCGCCTCAGG 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1136 gtgtacgcgggcaccagcaacctgcacagttgtcgtgagcgagcctcattgcgagatc 1195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GTGCAACTGCGGGAGCAGCCTTCTACTACAGGACGACGTCGCTGCGGTCAGCAGATC 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1196 atcatcaacagcaattacacagtagagagacgactatgacatcgccctcactgcggtg 1255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 ATCTGTGACCCACACAGTTCTACACCGCCAGATCGGAGACGACGTCGCCATCGCTGAG 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1256 tcaagaccctgacccctgtccgttcacatccaccctgttgcctcccatgcatggagac 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 GAGGAGCGGTGAAGGTCTCCAGCCATCCACACGTCACCCCTGCCCCCTCGCTCAGAG 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1316 acctttagcctcaatgagacgtctgtagcacaggtttggcaagaccagggagacagat 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 ACCTTCCCCCGGGAGTGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1376 gacaagacatccccctt---cctccggagaggtgcaggtcaatctcatcgactcaagaaa 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 CGCCTTCCACCCGCAATTTCTCTGAAGCAGGTGAAGTGTCCCATTAATGGAACACCAT 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1433 tg-----caatgactacttggctctatgacgttaaccttaccctcccaagatgtgtgct 1486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 TGTGACCGCAAAATACACACTTGGCGCTACACGGGAGACGACGTCGCCATCGCTGAG 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1487 ggggacctgtggtgggc-----agagactcctgccaggagagacgctgggggcct 1537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 GACATGCTGTGCGGGAGCAGCCTTCTACTACAGGACGACGTCGCCATCGTCCGTGAC 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1538 ctgtctgtgagcagaacacccctgtgtacctgtgcaggttaccacagctgggacagagc 1597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 CTGCTGTGCAAGGTGAATGTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1598 tgtgttcagagagaacaaacctgtgtgtacacacaaagtgcagaaagtcttccctgatt 1657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 TGTGCCAGCCCAACCGGCTTGCATCTACACCGCTGCTACCTACTACTTGGACTGGATC 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

; INFORMATION FOR SEQ ID NO: 13:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1137 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
US-08-978-404B-13

```

Query Match	6.9%;	Score 120.4;	DB 2;	Length 1137;
Best Local Similarity	52.7%;	Pred. No. 2.1e-20;		
Matches	396;	Conservative 0;	Mismatches 326;	Indels 30; Gaps 5;
QY	965	gggcggatctggagggcgctgctcgatagcaagtgcccttgcaagtgaagctcg	1024	
DB	81	GTGGGCAATGTCGGGGGTCAAGAGGCCCCAGGAGCAAGTGGCCCTGCAGGTGAGCGT	140	
QY	1025	cacttggcacc-----accacatctgtgaggcacgctcattgacgcccagttg	1075	
DB	141	AGAGTCACGGCCCATACTAGATGCACCTTCTGCGGGGCTCCCTCATCCACCCACAGTGG	200	
QY	1076	gtgctcaactgcgcccaactgcttcttgtagcccgggagaaagtccttgagggctgaag	1135	
DB	201	GTGCTGACCCAGCGCACTG---CGTGGGACCGGACGTCGAAGATCTGTGGCGCCCTCAGG	257	
QY	1136	gtgtacgcgggcacccgcaactcaccagttgctgaggaagcctcattgcccagatc	1195	
DB	258	GTGCAACTGCGGGAGCAGCACCTCTACTACCAGGACCAGTGCTGCGGGTCAGCAGGATC	317	
QY	1196	atcatcaacgcaattacacoga tgaaggagcaactatgacatgcgctcatcgctgt	1255	
DB	318	ATCGTGACCCACAGTTCTACACCGCCAGATCGGAGCGGACATCGCCCTGCTGGAGCTG	377	
QY	1256	tccaagccctgacctgtccgctcacatcacacctgtgctccccatgatggacag	1315	
DB	378	GAGGAGCGGTGAACGFTCCAGCCAGCTCCACAGGTACCCCTGCCCTTCGCTCAGAG	437	
QY	1316	acctttagctcaatgagacctgctggtatcacaggtcttggcaagacaggagacagat	1375	
DB	438	ACCTTCCCCCGGGATGCGTGTGGGTCTACTGGCTGGGCGATGTGGCAATGATGAG	497	
QY	1376	gacaagacatccccctt---cctcgggaggtgcaggtcaatctcatcgacttcaagaaa	1432	
DB	498	CGCCTCCACCGCCCATTTCTCTGAACGAGGTGAAGTGTCCCATTAATGGAACACCATT	557	
QY	1433	tg-----caatgactacttggtctatgacagttaccttaccccaagatgatgtgct	1486	
DB	558	TGTGACCCAAATACCACTTGGCGCTTACACGGGACGACGCTCCGCTCGTCCGTGAC	617	
QY	1487	ggggacctgctgggggc-----agagactcctgcaggagacagcgggggacct	1537	
DB	618	GACATGTGTGTGCGGGGAACACCCGGAGGGACTCATGCGGGGCGACTCGGAGGGGCC	677	
QY	1538	cttgtctgtgacagaacaaccgtgtaactggcaggttcacacgtggggcacaagc	1597	
DB	678	CTGGTGTGCAAGTGAATGACCTTGCTGCAGCGGGCGGTGTGTGACGTGGGGCGAGGGC	737	
QY	1598	tgtggccagagaacaacctggtgtgataccaaaagtgcacagaagttcttccctggatt	1657	
DB	738	TGTGCCACGCCAACCGGCTTGGCATCTACACCGTGTACCTTACTTCTGGACTGGATC	797	
QY	1658	tacagcaagatgagagcgagggtgcgattcag	1689	
DB	798	CACCACATATGCCCCAAAAAGCCGTGAGTCAG	829	

RESULT 14
US-09-016-366A-20
; Sequence 20, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu

```

1  TITLE OF INVENTION:  MAST CELL PROTEASE PEPTIDE
2
3  TITLE OF INVENTION:  INHIBITORS
4
5  NUMBER OF SEQUENCES:  65
6
7  CORRESPONDENCE ADDRESS:
8
9  ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
10
11 STREET:  600 Atlantic Avenue
12
13 CITY:  Boston
14
15 STATE:  MA
16
17 COUNTRY:  U.S.A.
18
19 ZIP:  02210-2211
20
21 COMPUTER READABLE FORM:
22
23 MEDIUM TYPE:  Diskette
24
25 COMPUTER:  IBM Compatible
26
27 OPERATING SYSTEM:  DOS
28
29 SOFTWARE:  FastSeq for Windows Version 2.0
30
31 CURRENT APPLICATION DATA:
32
33 APPLICATION NUMBER:  US/09/016,366A
34
35 FILING DATE:  January 30, 1998
36
37 CLASSIFICATION:  530
38
39 PRIORITY APPLICATION DATA:
40
41 APPLICATION NUMBER:  60/037,090
42
43 FILING DATE:  05-FEB-1997
44
45 ATTORNEY/AGENT INFORMATION:
46
47 NAME:  Plumer, Elizabeth R.
48
49 REGISTRATION NUMBER:  36,637
50
51 REFERENCE/DOCKET NUMBER:  B0801/7093
52
53 TELECOMMUNICATION INFORMATION:
54
55 TELEPHONE:  617-720-3500
56
57 TELEFAX:  617-720-2441
58
59 TELEX:
60
61 INFORMATION FOR SEQ ID NO:  20:
62
63 SEQUENCE CHARACTERISTICS:
64
65 LENGTH:  1128 base pairs
66
67 TYPE:  nucleic acid
68
69 STRANDEDNESS:  single
70
71 TOPOLOGY:  linear
72
73 MOLECULE TYPE:  CDNA
74
75 IS-09-016-366A-20

```

Query Match	6.8%;	Score 118.8;	DB 2;	Length 1128;
Best local Similarity	52.5%;	Pred. No. 5.3e-20;		
Matches 395;	Conservative 0;	Mismatches 327;	Indels 30;	Gaps 5;
Qy	965	gggcggatcgtggygaggggcgtggccctcgatagacaagtggccttggccaagtgaagtctg	1024	
Db	83	GTGGGCATCGTTGGGGGTTCAGGAGGCCCCAGGAGCAGTGGCCCTGCGAGGTGAGCCTG	142	
Qy	1025	catttcgggaccc-----accacaatctgttggaggcacgcctcaattgacgcgcagtgg	1075	
Db	143	AGAGTCCACGGCCCATACTGTGATCACTTCTGCGGGGGTCCCTCATCCACCCCCAGTGG	202	
Qy	1076	gtgctcaactgcgcgccactgcttcttcgtgaccgcgggagaaaggttccttgagggctgggaag	1135	
Db	203	GTGCTGACCGGACGCGCACTG---CGTGGGACCGGACGTCAGAGGATCTGGCCGCCCTCAGG	259	
Qy	1136	gtgtacgcgggcacacgaacactgcacagttgctgagggcagcctccattgcogagatc	1195	
Db	260	GTGCAACTGCGGGAGCAGCACTCTACTTACCAGGACCACAGCTGCTGCCGGTCAGCAGGATC	319	
Qy	1196	atcaatcaacagcaattacacagatgagaggagcaactatgacatgcgcctcatgcgctg	1255	
Db	320	ATCGTGACCCACAGTTCTTACACGCCCCAGATCGGAGCGGACATCGCCCTCTGCTGGAGCTG	379	
Qy	1256	tccaagccctgacccctgtccgcctcaatcacaccctgcttgcctcccccatgattggacag	1315	
Db	380	GAGGAGCGGTGAAGGTCTCCAGGCACGTCCACACGGTCAACCCTGCCCCCTGCCTCAGAG	439	
Qy	1316	acctttagctccaatgagacctgtgtgatacacagccttttggcaagaccaggagacagat	1375	
Db	440	ACCTTCCCCCCGGGATGCGCTGTGGTCTACTGGCTGGGGCGATGTGGACAAATGATGAG	499	
Qy	1376	qacaagacatccccctt---cctccggagagtgaggtccaatctcaatgcacttcaagaaa	1432	

```

Db 500 CGCTCCACCGCCATTCCCTCTGAAGCAGGTGAAGTCCCATTAATGGAAACACACATT 559
Qy 1433 tg-----caatgactacttggtctatgacagtttaccttaccctaccccaaggatgatgtgtct 1486
Db 560 TGTGACGCAAAATACCACTTGGCGCTACACGGGAGACGCTCCGCTCGTGAC 619
Qy 1487 ggggacctctggggg-----agagactctgcccaggagacagcgggggacct 1537
Db 620 GACATGTGTGTCGGGGAACACCGGAGGACTCATGCCAGGCGACTCCGGAGGGGCC 679
Qy 1538 ctctctgtgagcagaacaaccgtgtgtacctggcaggtgtcacagctgggggacaggc 1597
Db 680 CTGTGTGCAAGGTGAATGTCACCTGGCTGCAGGCGGCGTGGTCACTACTTGGACTGGATC 739
Qy 1598 tgtgcccagagaacaacctggtgtgtacaccaaagtgtacagaagttcttccctggatt 1657
Db 740 TGTGCCACGCCACCGCGCTGGCATCTACACCGCTGTACACCTACTTGGACTGGATC 799
Qy 1658 tacagcaagatggagagcgggtgcgattcag 1689
Db 800 CACCACATATGTCCCAAAAGCCGTGAGTCAG 831

```

RESULT 15

```

US-08-978-404B-15
; Sequence 15, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-15

```

Query Match 6.8%; Score 118.8; DB 2; Length 1128;
 Best Local Similarity 52.5%; Pred. No. 5.3e-20;
 Matches 395; Conservative 0; Mismatches 327; Indels 30; Gaps 5;

```

Qy 965 gggcgagatcgtggaggggcgctggccctcgatagcaagtgcccttggaagtgagtctg 1024
Db 83 GTGGGCATCTGTTGGGGTCTAGAGAGCCGCCAGAGCAAGTGGCCCTGTCAGTGCCTG 142
Qy 1025 cacttcgggaccc-----accacatctgtgaggcagcctcattgagcccgagtg 1075
Db 143 AGAGTCCACGGGCCCATCTGATGACATCTCTGCGGGGCTCCCTCATCCACCCCGAGTGG 202
Qy 1076 gtgtcactgcgcgccactgcttctgtgacccggggagaggtccttgagggctgggaag 1135
Db 203 GTGTGACCGCAGCGCAGCTG--CGTGGACCGGAGCGTCAAGGATCTTGCGCCGCTCAGG 259
Qy 1136 gtgtacgogggacacacacacacacacacacacacacacacacacacacacacacac 1195
Db 260 GTGCAACTGCGGGAGCAGCAGCTCTACTTACACGAGCAGCAGCTGCTGCCGCTCAGAGGATC 319
Qy 1196 atctcaacagcaattacacacacacacacacacacacacacacacacacacacacacac 1255
Db 320 ATCGTGCACCCACACAGTCTTACACCGCCCGAGATCGGAGCGGACATCGCCCTGCTGAGGCTG 379
Qy 1256 tccaagcccttgacccctgctccgctccacacacacacacacacacacacacacacacacac 1315
Db 380 GAGGAGCGGTGAAGGTCTCCAGCCACGTCACACGCTCACCCTGCCGCCCTGCCCTCAGAG 439
Qy 1316 accttagcctcaatgagacacacacacacacacacacacacacacacacacacacacacac 1375
Db 440 ACCTTCCCGCGGGATGCGCGTGTGGGTCTACTGCTGGGCGGATGTGGCAATGATGATGAG 499
Qy 1376 gacaagacatccccctt---cctccgggaggtgaggtcaatctcatcagacttcaagaaa 1432
Db 500 CGCCTCCACCGCCATTCTCTTGAAGCAGGTGAAGGTGCCCATTAATGGAAACACACATT 559
Qy 1433 tg-----caatgactacttggtctatgacacacacacacacacacacacacacacacacac 1486
Db 560 TGTGACGCAAAATACCACTTGGCGCTTACACGGGAGACGACGTCGCCGATCTGTCGGTGAC 619
Qy 1487 ggggacctctgtggggg-----agagactctgcccaggagacagcgggggacct 1537
Db 620 GACATGTGTGTCGGGGAACACCGGAGGACTCATGCCAGGCGACTCCGGAGGGGCC 679
Qy 1538 ctgtctgtgagcagaacaaccgtgtgtacctggcaggtgtcacagctgggggacaggc 1597
Db 680 CTGTGTGCAAGGTGAATGTCACCTGGCTGCGTGCAGGCGGCGTGGTCACTGTCGGGCGAGGGC 739
Qy 1598 tgtggccagagaacaacacacacacacacacacacacacacacacacacacacacacacac 1657
Db 740 TGTGCCACGCCCAACCGCGCTGGCATCTACACCGCTGTACACCGCTGTACTACTTGGACTGGATC 799
Qy 1658 tacagcaagatggagagcgggtgcgattcag 1689
Db 800 CACCACATATGTCCCAAAAGCCGTGAGTCAG 831

```

Search completed: August 19, 2002, 18:38:36
 Job time: 9074 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 21:23:43 ; Search time 292.6 Seconds
(without alignments)
10256.891 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctgagagaccatgagagagg.....ggctgctgtgactcgagaaa 1748

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1678	96.0	1689	22	Human transmembran
2	1663	95.1	2067	22	Human transmembran
3	1253	71.7	1314	22	Human SER6 nucleot
c 4	1253	71.7	1314	22	Nucleotide sequenc
5	1049	60.0	2192	22	Human cDNA sequenc
c 6	1019	58.3	1078	22	Nucleotide sequenc
7	782	44.7	1222	21	Human serine prote
8	257	14.7	852	22	Human cDNA clone (
9	250	14.3	1212	23	DNA encoding novel

10	156	8.9	926	22	Human reproductive
11	152	8.7	450	22	Novel human diagno
12	146	8.4	705	23	DNA encoding novel
13	145	8.3	926	22	Human reproductive
c 14	29	1.7	29	22	Human endotheliase
c 15	29	1.7	29	22	Human endotheliase
16	27	1.5	27	22	Human endotheliase
17	27	1.5	27	22	Human endotheliase
18	26	1.5	26	22	Human endotheliase
19	25	1.4	25	22	Human endotheliase
c 20	25	1.4	36	22	Human endotheliase
21	24	1.4	39	22	Human endotheliase
22	24	1.4	509	23	DNA encoding novel
c 23	24	1.4	969	22	Human Stratum corn
24	24	1.4	986	16	Human stratum corn
c 25	24	1.4	1089	17	Human amyloid prec
26	23	1.3	23	22	SER7 gene specific
27	22	1.3	22	22	Human endotheliase
28	22	1.3	1482	23	Drosophila melanog
29	22	1.3	3630	23	SER7 gene specific
c 30	21	1.2	21	22	Viral encoded sema
31	21	1.2	21	22	Human VESPR cDNA.
32	21	1.2	4707	20	Viral encoded sema
33	21	1.2	4707	21	Human VESPR cDNA.
34	21	1.2	4707	22	Viral encoded sema
35	21	1.2	4707	22	Human VESPR cDNA.
36	21	1.2	4707	22	Nucleotide sequenc
37	21	1.2	4707	22	CDNA encoding mous
c 38	20	1.1	251	21	CDNA encoding mous
39	20	1.1	717	21	CDNA encoding mous
40	20	1.1	1685	21	CDNA encoding mous
41	20	1.1	2068	21	CDNA encoding mous
42	20	1.1	2070	21	Human pollinosis-a
43	19	1.1	631	21	Human pollinosis-a
44	19	1.1	1030	21	Human pollinosis-a
45	19	1.1	1034	21	CDNA encoding mous

ALIGNMENTS

RESULT 1

AAD05796

ID AAD05796 standard; DNA; 1689 BP.

AC AAD05796;

XX

DT 31-JUL-2001 (first entry)

XX Human transmembrane serine protease (Endotheliase 2-S) DNA.

XX Human; endotheliase 2-S; protease domain; cytosolic; vulnary; wound;
KW nontropic; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
KW liver cirrhosis; osteoradionecrosis; systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;
KW transmembrane serine protease; ds.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 1..1689

FT /tag= a

FT /product= "Human endotheliase 2-S protein"

XX WO200136604-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-US31803.

XX 18-NOV-1999; 99US-0166391.
 PR 22-SEP-2000; 2000US-0234840.
 XX
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 XX Madison EL, Ong EO;
 PI
 XX WPI; 2001-336001/35.
 DR P-PSDB; AAE01943.
 DR
 XX New nucleic acid encoding a protein comprising endotheliase activity
 PT useful in the prevention and treatment of e.g. vascular malformations,
 PT cardiovascular disorders, and chronic inflammatory disease -
 XX
 XX Example 2; Page 135-137; 152pp; English.
 PS
 XX The present DNA sequence encodes human short form transmembrane serine
 CC protease (Endotheliase 2-S) protein.
 CC The invention relates to an endotheliase protein, endotheliase protease
 CC domain and their corresponding nucleic acid molecules. An endotheliase
 CC protein or protease domain of it is useful for the treatment and
 CC diagnosis of disorders associated with aberrant angiogenesis or undesired
 CC neovascularisation. The undesired angiogenesis is associated with
 CC disorders selected from solid neoplasm, vascular malformations and
 CC cardiovascular disorders such as angiofibroma, angiolipoma,
 CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
 CC malformations, haemangiomatosis and vascular adhesions, dyschondroplasia
 CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
 CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
 CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
 CC hemophilic joints, inflammatory bowel disease, nonhealing fractures,
 CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
 CC granulations-burns, hypertrophic scars, liver cirrhosis,
 CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
 CC systemic sclerosis and aberrant wound repairs, circulatory disorders
 CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
 CC dyomeioty, sclerodactyly and teangiectasis, dermatological disorders
 CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
 CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
 CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
 CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
 CC caused by ocular neovascular disease, corneal graft neovascularisation,
 CC macular degeneration, retinopathy of prematurity, retrolental
 CC fibroplasia and corneal neovascularisation. The nucleic acids of the
 CC invention are also used in gene therapy. The invention also provides
 CC method for screening compounds that modulate angiogenesis.
 XX
 SQ Sequence 1689 BP; 367 A; 529 C; 470 G; 323 T; 0 other;

Query Match 96.0%; Score 1678; DB 22; Length 1689;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 atggagaggacagccaggaatgcattccagcaagaacaccttcagctgaggaatct 70
 DB 1 atggagaggacagccaggaatgcattccagcaagaacaccttcagctgaggaatct 60
 QY 71 ccagccaggacatctcagctggacacctccagccggcgatctccagccaggcatct 130
 DB 61 ccagccaggacatctcagctggacacctccagccggcgatctccagccaggcatct 120
 QY 131 ccagccaggacatctcagctggacacctccagccggcgatctccagccaggcatct 190
 DB 121 ccagccaggacatctcagctggacacctccagccggcgatctccagccaggcatct 180
 QY 191 ccagctggtacacctccagccggcgatctccagccggcgatctccagccaggcatct 250
 DB 181 ccagctggtacacctccagccggcgatctccagccggcgatctccagccaggcatct 240
 QY 251 ccagccgggacatctccgctctggacatcattccaggtccctcatccgagcaggtcatca 310
 DB 1321 gagacctgctggtacacagcgttttggcaagaccaggagagacagatgacaagacatcccc

Db 241 ccagccgggacatctccggtctctggatactttccagggtctctccatccgagcaggtcatca 300
 QY 311 tccgcccaggtcagctcgttgacaacctcccccaaccagagtgtagcttggtagagaaca 370
 Db 301 tccgcccaggtcagctcgttgacaacctcccccaaccagagtgtagcttggtagagaaca 360
 QY 371 ccagtgggggtgtacccatccgatactctctgcagaggtcagccaccagcaaccaggggcc 430
 Db 361 ccagtgggggtgtacccatccgatactctctgcagaggtcagccaccagcaaccaggggcc 420
 QY 431 accagggagagccaggtacagcctgcccaagtccaaccttcagcgggagggccagagaacag 490
 Db 421 accagggagagccaggtacagcctgcccaagtccaaccttcagcgggagggccagagaacag 480
 QY 491 ctaccgctcagcgggtcgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 550
 Db 481 ctaccgctcagcgggtcgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 540
 QY 551 ttcagttctggcagggccacacagggatcaggtacaagagagagagagagagagctgtccc 610
 Db 541 ttcagttctggcagggccacacagggatcaggtacaagagagagagagagagagctgtccc 600
 QY 611 aagcacgcttctcgtcgtgacggtggtggtgactgcaagctgaagagtgacgagctgggc 670
 Db 601 aagcacgcttctcgtcgtgacggtggtggtgactgcaagctgaagagtgacgagctgggc 660
 QY 671 tgcgtgaggtttgactggggacaagtctcgtctaaatctactctcctcctcctcctcctcctc 730
 Db 661 tgcgtgaggtttgactggggacaagtctcgtctaaatctactctcctcctcctcctcctcctc 720
 QY 731 tgccttccatctgtagcagcaactggaatgactcctcactcagagaagaactgccagcag 790
 Db 721 tgccttccatctgtagcagcaactggaatgactcctcactcagagaagaactgccagcag 780
 QY 791 ctgggttccagagtgctccaccgacaacccaggtgtccacacagggattttgccaaacagc 850
 Db 781 ctgggttccagagtgctccaccgacaacccaggtgtccacacagggattttgccaaacagc 840
 QY 851 tctcaatcttgagatacaactccacatccaggaagcctccacaggtctgaatgcct 910
 Db 841 tctcaatcttgagatacaactccacatccaggaagcctccacaggtctgaatgcct 900
 QY 911 tccacggtatctcctccaggttccactcggactgagggcagggccatgacggcgcg 970
 Db 901 tccacggtatctcctccaggttccactcggactgagggcagggccatgacggcgcg 960
 QY 971 atcgtggagggcgctggcctcgatagcaagtggccttggaagtgaagtgcacttc 1030
 Db 961 atcgtggagggcgctggcctcgatagcaagtggccttggaagtgaagtgcacttc 1020
 QY 1031 ggacacccacatctgtggaggcagcgtcattgacgcccagtggtgtcactgcgcgc 1090
 Db 1021 ggacacccacatctgtggaggcagcgtcattgacgcccagtggtgtcactgcgcgc 1080
 QY 1091 cactgcttctcgtgacccgggagaaggtcctgagggcgtggaaggtgtacgcgggcacc 1150
 Db 1081 cactgcttctcgtgacccgggagaaggtcctgagggcgtggaaggtgtacgcgggcacc 1140
 QY 1151 agcaacctgaccaggttgcctgaggagcctccatgcccagatcatcatacagcaat 1210
 Db 1141 agcaacctgaccaggttgcctgaggagcctccatgcccagatcatcatacagcaat 1200
 QY 1211 tacacgagtagagagagcactatgacatgcgctcctcctcctcctcctcctcctcctcctc 1270
 Db 1201 tacacgagtagagagagcactatgacatgcgctcctcctcctcctcctcctcctcctcctc 1260
 QY 1271 ctgctcgtcacatccacctgcttgcctcccatgcatgagacagaccttagcctcaat 1330
 Db 1261 ctgctcgtcacatccacctgcttgcctcccatgcatgagacagaccttagcctcaat 1320
 QY 1331 gagacctgctggatcacagcgttttggcaagaccaggagagacagatgacaagacatcccc 1390
 Db 1321 gagacctgctggatcacagcgttttggcaagaccaggagagacagatgacaagacatcccc 1380

QY 1391 ttctccggaggtgaggtcgaatctctcatctcgaatgaaatgcaatgactactgtgtc 1450
 |||||
 Db 1381 ttctccggaggtgaggtcgaatctctcatctcgaatgaaatgcaatgactactgtgtc 1440
 |||||
 QY 1451 tatgacagttacattcccccaagatgattgtgctgggacattctgtggggcagagac 1510
 |||||
 Db 1441 tatgacagttacattcccccaagatgattgtgctgggacattctgtggggcagagac 1500
 |||||
 QY 1511 ttctgcaggagagacagcgggggcctcttctgtgagcagacaacccgtgtaccctg 1570
 |||||
 Db 1501 ttctgcaggagagacagcgggggcctcttctgtgagcagacaacccgtgtaccctg 1560
 |||||
 QY 1571 gcagggttcaccagctgggggcacaggtgtggccagagaaacaaacctgtgtgtacac 1630
 |||||
 Db 1561 gcagggttcaccagctgggggcacaggtgtggccagagaaacaaacctgtgtgtacac 1620
 |||||
 QY 1631 aaagtacagaattcttccctgattacagcaagatggagagcgaggtgcattca 1688
 |||||
 Db 1621 aaagtacagaattcttccctgattacagcaagatggagagcgaggtgcattca 1678
 |||||

RESULT 2

AAD05797

ID AAD05797 standard; DNA; 2067 BP.

AC AAD05797;

DT 31-JUL-2001 (first entry)

DE Human transmembrane serine protease (Endotheliase 2-L) DNA.

XX Human; endotheliase 2-L; protease domain; cytotstatic; vulnerary; wound;
 KW neotropic; periodontitis; dermatological disorder; gene therapy; scar;
 KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
 KW chronic inflammatory disease; ocular disorder; circulatory disorder;
 KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
 KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;
 KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
 KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;
 KW transmembrane serine protease; ds.

XX Homo sapiens.

FH Key Location/Qualifiers

FT 1..2067

FT /*tag= a

FT /product= "Human endotheliase 2-L protein"

XX WO200136604-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-US31803.

XX 18-NOV-1999; 99US-0166391.

XX 22-SEP-2000; 2000US-0234840.

XX (CORV-) CORVAS INT INC.

XX Madison EL, Ong EO;

XX WPI; 2001-336001/35.

XX P-PSDB; AAE01944.

XX New nucleic acid encoding a protein comprising endotheliase activity
 PT useful in the prevention and treatment of e.g. vascular malformations,
 PT cardiovascular disorders, and chronic inflammatory disease -
 XX Example 2; Page 139-142; 152pp; English.

XX The present DNA sequence encodes human long form transmembrane serine

XX protease (Endotheliase 2-L) protein.

CC

CC The invention relates to an endotheliase protein, endotheliase protease
 CC domain and their corresponding nucleic acid molecules. An endotheliase
 CC protein or protease domain of it is useful for the treatment and
 CC diagnosis of disorders associated with aberrant angiogenesis or undesired
 CC neovascularisation. The undesired angiogenesis is associated with
 CC disorders selected from solid neoplasm, vascular malformations and
 CC cardiovascular disorders such as angiofibroma, angiolipoma,
 CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
 CC malformations, haemangiomas and vascular adhesions, dyschondroplasia
 CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
 CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
 CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
 CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
 CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
 CC granulations-burns, hypertrophic scars, liver cirrhosis,
 CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
 CC systemic sclerosis and aberrant wound repairs, circulatory disorders
 CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
 CC dysmotility, sclerodactyly and teangiectasis, dermatological disorders
 CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
 CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
 CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
 CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
 CC caused by ocular neovascular disease, corneal graft neovascularisation,
 CC macular degeneration, retinopathy of prematurity, retrolental
 CC fibroplasia and corneal neovascularisation. The nucleic acids of the
 CC invention are also used in gene therapy. The invention also provides
 CC method for screening compounds that modulate angiogenesis.

XX Sequence 2067 BP; 463 A; 604 C; 540 G; 460 T; 0 other;

SQ

Query Match 95.1%; Score 1663; DB 22; Length 2067;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 atggagggagacgacggggaatgcattccagcagaagaacaccttcagctggagcatct 70
 |||||
 Db 1 atggagggagacgacggggaatgcattccagcagaagaacaccttcagctggagcatct 60
 |||||

QY 71 caagccaggcatctccagctggggaacacctccagggcgccatctccagccaggcatct 130
 |||||
 Db 61 ccagccaggcatctccagctggggaacacctccagggcgccatctccagccaggcatct 120
 |||||

QY 131 ccagccaggcatctccagctggggaacacctccagggcgccatctccagccaggcatct 190
 |||||
 Db 121 ccagccaggcatctccagctggggaacacctccagggcgccatctccagccaggcatct 180
 |||||

QY 191 ccagctggtacacctccagggcgccatctccagggcgccatctccagccaggcatct 250
 |||||
 Db 181 ccagctggtacacctccagggcgccatctccagggcgccatctccagccaggcatct 240
 |||||

QY 251 ccagccgggcatctccggctctggcatacatttcaggttcctccagccaggcatctca 310
 |||||
 Db 241 ccagccgggcatctccggctctggcatacatttcaggttcctccagccaggcatctca 300
 |||||

QY 311 tccgcccaggtcagctcgtgtgacaacctccccaacagagtgacctgttagagaaca 370
 |||||
 Db 301 tccgcccaggtcagctcgtgtgacaacctccccaacagagtgacctgttagagaaca 360
 |||||

QY 371 ccagtgggggtgtaccatccgatcatctctgccaggtccaggtccagcagaacaggcc 430
 |||||
 Db 361 ccagtgggggtgtaccatccgatcatctctgccaggtccaggtccagcagaacaggcc 420
 |||||

QY 431 accagggagagccaggtacagctcgtcccaagttcacctgggggagggccagagagcag 490
 |||||
 Db 421 accagggagagccaggtacagctcgtcccaagttcacctgggggagggccagagagcag 480
 |||||

QY 491 ctaccctcatcggtgctgctccctccctcattgctcctggtgttctcgtcatcctc 550
 |||||
 Db 481 ctaccctcatcggtgctgctccctccctcattgctcctggtgttctcgtcatcctc 540
 |||||

QY 551 ttccagttctggcaggggccacacagggatcaggtacaaagagcagaggagagctgcc 610
 |||||

Db 541 ttccagttctgagggccacacagggatcaggtacagagagcagagagagctgtccc 600
Qy 611 aagcacgctgtctcgtgacggggtggtgactgcaagctgaagagtacgagctggc 670
Db 601 aagcacgctgtctcgtgacggggtggtgactgcaagctgaagagtacgagctggc 660
Qy 671 tgcgtgaggttgaactgggcaagctctctcttaaaatctactctggtctccatcag 730
Db 661 tgcgtgaggttgaactgggcaagctctctcttaaaatctactctggtctccatcag 720
Qy 731 tggcttccatctgtacagcaactgaatgactcttactcagagaagactgccagcag 790
Db 721 tggcttccatctgtacagcaactgaatgactcttactcagagaagactgccagcag 780
Qy 791 ctgggttccagagtgctccacggacacagaggtgtccacagagatttgcacacgc 850
Db 781 ctgggttccagagtgctccacggacacagaggtgtccacagagatttgcacacgc 840
Qy 851 ttctcaatcttgagatacaactccaccatccaggaagcgtccacaggttctgaatgcct 910
Db 841 ttctcaatcttgagatacaactccaccatccaggaagcgtccacaggttctgaatgcct 900
Qy 911 tccacggttatctccctccaggttccactcggagactgagggccatgacggcg 970
Db 901 tccacggttatctccctccaggttccactcggagactgagggccatgacggcg 960
Qy 971 atcgtggagggcgctggcctcgatagcaagtggccttggaagtgtctgcacttc 1030
Db 961 atcgtggagggcgctggcctcgatagcaagtggccttggaagtgtctgcacttc 1020
Qy 1031 ggcaccaccacatctctgaggcagctcattgacgcccagtggtgtcactgcgc 1090
Db 1021 ggcaccaccacatctctgaggcagctcattgacgcccagtggtgtcactgcgc 1080
Qy 1091 cactgtcttctgacccggagaaagtctcggaggtcggaggtgtacggggcacc 1150
Db 1081 cactgtcttctgacccggagaaagtctcggaggtcggaggtgtacggggcacc 1140
Qy 1151 agcaactgcaccagtctcgtgagcagcctccattccgagatcatcaacagcaat 1210
Db 1141 agcaactgcaccagtctcgtgagcagcctccattccgagatcatcaacagcaat 1200
Qy 1211 tacacgatgaggagagcactatgacatcgccctcgtcgtgtccaaagccctgacc 1270
Db 1201 tacacgatgaggagagcactatgacatcgccctcgtcgtgtccaaagccctgacc 1260
Qy 1271 ctgtccgctcacatccaccctgtgtcgtcccatgcatgacagacacttagcctcaat 1330
Db 1261 ctgtccgctcacatccaccctgtgtcgtcccatgcatgacagacacttagcctcaat 1320
Qy 1331 gagacctgctggatcacaggctttggaagaccagggagacagatgacaagacatcccc 1390
Db 1321 gagacctgctggatcacaggctttggaagaccagggagacagatgacaagacatcccc 1380
Qy 1391 ttctccggaggtgaggtgcaatctcatcgacttcaagaaatgcaatgactactgttc 1450
Db 1381 ttctccggaggtgaggtgcaatctcatcgacttcaagaaatgcaatgactactgttc 1440
Qy 1451 tatgacagttacattaccccaagatgattgtgtcgtgggacacttctgtgggagagac 1510
Db 1441 tatgacagttacattaccccaagatgattgtgtcgtgggacacttctgtgggagagac 1500
Qy 1511 tcttcgagggagacagcggggcctctgtgtgtgagcagacaacacctgtgtactctg 1570
Db 1501 tcttcgagggagacagcggggcctctgtgtgtgagcagacaacacctgtgtactctg 1560
Qy 1571 gcaggtgtaccagctggggcagagcgtgtggccagagaacaaacctgtgtgtacacc 1630
Db 1561 gcaggtgtaccagctggggcagagcgtgtggccagagaacaaacctgtgtgtacacc 1620
Qy 1631 aaagtacagagttcttccctggattacagcaagatggaga 1673

Db 1621 aaagtacagagttcttccctgattacagcaagatggaga 1663

RESULT 3

AAF83971 standard; DNA; 1314 BP.

AC AAF83971;

DT 06-AUG-2001 (first entry)

XX Human SER6 nucleotide sequence.

DE Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
KW cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
KW immunosuppressive; antidiabetic; antifibrinolytic; SER6; ds.

OS Homo sapiens.

XX Key Location/Qualifiers
CDS 1..1266
FT /*tag= a
FT /product= "SER6"

XX :WO200136645-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-US31744.

XX 17-NOV-1999; 99US-0165986.

XX 09-FEB-2000; 2000US-0181347.

XX 03-APR-2000; 2000US-0194195.

XX 05-APR-2000; 2000US-0194839.

XX 07-APR-2000; 2000US-0195637.

XX 13-APR-2000; 2000US-0197080.

XX 15-SEP-2000; 2000US-0232877.

XX 16-NOV-2000; 2000US-0165986.

XX (CURA-) CURAGEN CORP.

XX Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Taupier K, Rastelli L, Herrmann JL;

XX WPI; 2001-336006/35.

XX P-PSDB; AAB85042.

XX New isolated SERX polynucleotides and polypeptides related to the
PT serine/threonine kinase family of proteins, useful for determining and
PT treating diseases related to altered levels of the polypeptide -

XX Claim 9; Page 33; 127pp; English.

XX The invention relates to serine/threonine kinase-like proteins, termed
CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
CC via a recombinant expression vector in a host cell in gene therapy
CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
CC polypeptides, polynucleotides, antibodies can also be used to treat or
CC prevent a pathology associated with SERX. Disorders associated with
CC aberrant SERX expression or activity include breast or testicular cancer,
CC cardiovascular defect, autoimmune disorders and asthma. Disorders
CC associated with aberrant SERX-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
CC represents the nucleotide sequence of human SER6.

XX Sequence 1314 BP; 288 A; 384 C; 376 G; 266 T; 0 other;

Query Match 71.7%;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 ggagagccaggttacgagcctgccaaagttaacctgctgagggagggccagagcagctacc 495
Db 3 ggagagccaggttacgagcctgccaaagttaacctgctgagggagggccagagcagctacc 62
QY 496 gctcatcggttgctgctctctctcattgccttggttgccttgcctatcatctcttccaa 555
Db 63 gctcatcggttgctgctctctcattgccttggttgccttgcctatcatctcttccaa 122
QY 556 gtcttgccagggccacacagggatcagggtacaaggagacagagagagctgtcccaagca 615
Db 123 gtcttgccagggccacacagggatcagggtacaaggagacagagagagctgtcccaagca 182
QY 616 cgtgttcctgtgacggggttggtggactgcaagctgaagagtgacagctgggctgcgt 675
Db 183 cgtgttcctgtgacggggttggtggactgcaagctgaagagtgacagctgggctgcgt 242
QY 676 gaggtttgactggacaagtctctgtcttaaaatactctggttgcctcccatcagtggt 735
Db 243 gaggtttgactggacaagtctctgtcttaaaatactctggttgcctcccatcagtggt 302
QY 736 tcccatctgtagcagaactggaatgactctactcactcagagaagacctgccagcagctggg 795
Db 303 tcccatctgtagcagaactggaatgactctactcactcagagaagacctgccagcagctggg 362
QY 796 ttctcgaggtgctcaocgggacaaccagggttgcctcagagggatttgcctcagcagcttctc 855
Db 363 ttctcgaggtgctcaocgggacaaccagggttgcctcagagggatttgcctcagcagcttctc 422
QY 856 aatcttgagatacaactcccaatccaggaagacctccacagcttgaatgccttccca 915
Db 423 aatcttgagatacaactcccaatccaggaagacctccacagcttgaatgccttccca 482
QY 916 gcggtatatctctccctcagtgctccactgcgactgcgagggccatgacccggcgatcgt 975
Db 483 gcggtatatctctccctcagtgctccactgcgactgcgagggccatgacccggcgatcgt 542
QY 976 ggagggggcgctggctcgagtagcaagtggcccttgcaagtgcacttcagctcgacac 1035
Db 543 ggagggggcgctggctcgagtagcaagtggcccttgcaagtgcacttcagctcgacac 602
QY 1036 caccacatctgtaggagcagctcaattgacgcccagtggtgctcactgcgcccactg 1095
Db 603 caccacatctgtaggagcagctcaattgacgcccagtggtgctcactgcgcccactg 662
QY 1096 ctcttcctgtagcccgagagagctgctggaggctggaagtgtagcgggacccagca 1155
Db 663 ctcttcctgtagcccgagagagctgctggaggctggaagtgtagcgggacccagca 722
QY 1156 cctgcaccaggttgcctgagggagcctccattgcccagatcatcatcaacagcaattacac 1215
Db 723 cctgcaccaggttgcctgagggagcctccattgcccagatcatcatcaacagcaattacac 782
QY 1216 cagtagggagagcactatgacatccctccatgcgctgctccaaagccctgacctgtc 1275
Db 783 cagtagggagagcactatgacatccctccatgcgctgctccaaagccctgacctgtc 842
QY 1276 cgctcatcatccctctgcttccctcccatgcatggagacagacctttagctcctaatgagac 1335
Db 843 cgctcatcatccctctgcttccctcccatgcatggagacagacctttagctcctaatgagac 902
QY 1336 ctgctggaatacagggttggaagaccaggaggagacagatgacaagacatcccccttcc 1395
Db 903 ctgctggaatacagggttggaagaccaggaggagacagatgacaagacatcccccttcc 962
QY 1396 ccggagaggtgcaggtcaatctcatcgaactcaagaaatgcaatgactacttggtctatga 1455
Db 963 ccgggaggggcaggtcaatctcatcgaactcaagaaatgcaatgactacttggtctatga 1022
QY 1456 cagttaccttaccacagagatgctgctggggagcctctgctggggcagagactcctg 1515
Db 1023 cagttaccttaccacagagatgctgctggggagcctctgctggggcagagactcctg 1082
QY 1516 ccagggagagacggggggcctcttctgtgtgagcagaacacccgctggtacctggcag 1575

Db 1083 ccagggagacagcgggggctcttctgtgtgagcagaacacccgctggtacctggcag 1142
QY 1576 tgtcacagctggtggccacagcgtgtggccagagaaaaaacacctggtgtgtacacaaagt 1635
Db 1143 tgtcacagctggtggccacagcgtgtggccagagaaaaaacacctggtgtgtacacaaagt 1202
QY 1636 gacagaagtcttccctcctggtattacagaagatggagcagcgtgcgattca 1688
Db 1203 gacagaagtcttccctcctggtattacagaagatggagcagcgtgcgattca 1255

RESULT 4

AAF83973/c

ID AAF83973 standard; DNA; 1314 BP.

AC AAF83973;

DT 06-AUG-2001 (first entry)

DE Nucleotide sequence of SER8, a partial reverse complement of SER6.

XX Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
KW cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
KW immunosuppressive; antiasthmatic; antifibrinolytic; SER6; SER8; ss.
XX Homo sapiens.

OS Homo sapiens.

XX WO200135645-A2.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-US31744.

XX 17-NOV-1999; 99US-0165986.

PR 09-FEB-2000; 2000US-0181347.

PR 03-APR-2000; 2000US-0194195.

PR 05-APR-2000; 2000US-0194839.

PR 07-APR-2000; 2000US-0195637.

PR 13-APR-2000; 2000US-0197080.

PR 15-SEP-2000; 2000US-0232877.

PR 16-NOV-2000; 2000US-0165986.

XX (CURA-) CURAGEN CORP.

XX Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;

PI Taupier R, Rastelli L, Herrmann JL;

XX WPI; 2001-336006/35.

XX New isolated SERX polynucleotides and polypeptides related to the

PT serine/threonine kinase family of proteins, useful for determining and

PT treating diseases related to altered levels of the polypeptide -

XX Claim 9; Page 38; 127pp; English.

XX The invention relates to serine/threonine kinase-like proteins, termed

CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.

CC via a recombinant expression vector in a host cell in gene therapy

CC applications), to detect SERX mRNA (e.g. in a biological sample) or a

CC genetic lesion in a SERX gene and modulate SERX activity. The SERX

CC polypeptides, polynucleotides, antibodies can also be used to treat or

CC prevent a pathology associated with SERX. Disorders associated with

CC aberrant SERX expression or activity include breast or testicular cancer,

CC cardiovascular defect, autoimmune disorders and asthma. Disorders

CC associated with aberrant SERX-8 expression include, blood clotting

CC hematopoietic and tumor related disorders. The present sequence

CC represents the nucleotide sequence of human SER8, a partial reverse

CC complement of SER6.

XX Sequence 1314 BP; 266 A; 376 C; 384 G; 288 T; 0 other;

SQ

Query Match				71.78;	Score 1253;	DB 22;	Length 1314;
Best Local Similarity				100.0%;	Pred. No. 0;		
Matches 1253;				Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	436	ggagagccaggtacagagcctgcccaggttcacactggcgagggagccagagcagctacc	495				
Db	1312	GGAGAGCCAGGTACGAGCCTGCCAAGTTCACCTGGCGGAGGGCCAGAGAGCAGCTACC	1253				
Qy	496	gctcatcggtgctgtccca	555				
Db	1252	GCTCATCGGGTGGCTGCTCA	1193				
Qy	556	gtcttgccagggccacacagagatcaggtacagagagagagagagagagagagagagagagca	615				
Db	1192	GTTCTGGCAGGGCCACACAGAGGATCAGGTACAGAGAGAGAGAGAGCTGTCCCAAGCA	1133				
Qy	616	cgctgttcgtgtgacgggtggtggactgcaagctgaagctgaagctgacagctgggctgcgt	675				
Db	1132	CGCTGTTGCTGTGACGGGTGGTGGACTGCAAGCTGAAGAGTGACGAGCTGGCTGGCT	1073				
Qy	676	gaggttgactgggacaggtctctgcttaaatctactctctggttctctctctctctctctctc	735				
Db	1072	GAGGTTTGACTGGGACAAAGTCTGCTTAAATCTACTCTGGTCTCTCCCATCAGTGGCT	1013				
Qy	736	tcccatctgtagcagcaactggaatgactctactctactcagagagacctgcagagagctggg	795				
Db	1012	TCCCATCTGTATGACAACTCCGAATGACTCTACTCAGAGAAAGACCTGCCACGAGCTGGG	953				
Qy	796	tttcgagagctctaccggcaaacagaggttgccacagggatgttgcacacagctctctc	855				
Db	952	TTTCGAGAGTCTCTACCGGACAAACGAGGTGGCCACAGGGATTTTGCCAAACAGCTTCTC	893				
Qy	856	aatcttgagatacaactccaccatccagggaaagcctccacaggtctgaatgcccttccca	915				
Db	892	AATCTTGAGATACAACTCCACCATCCAGGAAAGCTCCACAGGCTGTAATGCCCTTCCCA	833				
Qy	916	gcggtatctccctccaagtgttccactgcggactgagggccatgacccggggcgatcgt	975				
Db	832	CGGATATATCTCCCTCCAGTCTTCCCACTGCGGACTGAGGGCCATGACCGGGCGGATCGT	773				
Qy	976	gggaggggcgctggcctcgatagcaagtggccttggcgaagtggactgcacttcggcac	1035				
Db	772	GGGAGGGCGCTGGCCTCGGATACGAAGTGCCCTTGGCAAGTGAGCTGCGACTTCGGCAC	713				
Qy	1036	caccacatctgttgaggacagctcatgtgacgcccagtggtgtcactgcgcgccactg	1095				
Db	712	CACCCACATCTGTGAGGACAGCTCATTTGACGCCAGTGGGTGCTCACTGCCGCCCACTG	653				
Qy	1096	cttctctgacccgggagagaaggtccctggaggctggaaggtgtacggggcaccagcaa	1155				
Db	652	CTTCTTGTTGACCCGGGAGAGGTCTCTGGAGGGCTGGAAGGTGTACCGGGCACCCAGCAA	593				
Qy	1156	cctgcaccaggtgctgaggcagctccattgcggagatcatcatcaacagcaattacac	1215				
Db	592	CCTGCACCAGTTGGCTGAGGCAGGCTCCATTTGCCAGATCATCATACAGCAATTACAC	533				
Qy	1216	cgatgaggagacactatgacatgcctctcatgctgggtgtccaaagccctgacccctgtc	1275				
Db	532	CGATGAGGAGGACGACTATGACATCGCCCTCATGCGGCTGTCCAAAGCCCTGACCCCTGTC	473				
Qy	1276	cgctcacatcacccctgctgtctccctccatgcatggacagacaccttagcctcaatgagac	1335				
Db	472	CGCTCACATCCACCCTGCTGCTCCCTCCCATGATGAGAGACCTTTAGCCTCAATGAGAC	413				
Qy	1336	ctgctggtatcacaggtttggcaagaccagggagagacagatgacagacatccccttccct	1395				
Db	412	CTGCTGGATCACAGGCTTTGGCAAGACCAGGAGACAGATGACAAAGACATCCCCCTTCCCT	353				
Qy	1396	ccgggaggtgcaggttcaatctcatcgacttcaagaataatgcaatgactacttggctctatga	1455				
Db	352	CCGGAGGTGCGAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTACTTGGTCTATGA	293				
Qy	1456	cagttaccttaccaccaagatgatgtgtgctggggaccttctgtggggcgagagactcgtg	1515				

Db	292	CAGTTACCTTACCCCAAGGATGATGTGTGCTGGGACCTTCTGTTGGGCAGAGACTCTGTG	233
Qy	1516	ccagggagacagcgggggacctcttctgtgtgagcagaacacccgctggtgtacctggcagg	1575
Db	232	CCAGGGAGACAGCGGGGGCCTCTTGTCTGTGTGAGCAGAAACACCGCTGTTACCTGGCAGG	173
Qy	1576	tgtcacacagctggggcagcagcgtgtgcccagagagaaaaaacctggtgtgtacaccaaagt	1635
Db	172	TGTCAACAGCTGTGGGCAACAGGCTGTGGCCAGAGAAACAAACCTGTGTGTACCCAAAGT	113
Qy	1636	gacagaagttctccctggatttacagcaaatgagagcaggtggcattca	1688
Db	112	GACAGAAGTTCTTCCCTGGATTTACAGCAAGATGGAGAGCGAGGTGCGATTCA	60

RESULT 5
AAH14850
ID AAH14850 standard; cDNA; 2192 BP.
XX
AC AAH14850;
XX
DT 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:12681.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
XX
PN EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
PS Claim 8; SEQ ID 12681; 2537pp + CD ROM; English.
XX

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length

Query Match		58.3%; Score 1019; DB 22; Length 1078;
Best Local Similarity		100.0%; Pred. No. 0;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	670	ctgggtgaggtttgactgggagcaagctctctcttaaaatctactctgggtctcccatca 729
Db	1078	CTGGGTGAGGTTGACTGGGCAAGTCTCTCTTTAAATCTACTCTGGGTCTCCCATCA 1019
Qy	730	gtggcttcccatctgtagcagcaactgaatgactctctactcagagaagacctccagca 789
Db	1018	GTGGCTTCCCATCTGTAGCACCACTTGAATGACTCTCTACTCTCAGAGAAGACTCCACGCA 959
Qy	790	gctgggttcagagtgctcaccggacaacacagggttgccacacagggtattttgccaacag 849
Db	958	GCTGGGTTTCGAGAGTGCTCACCAGCAACAGAGGTTGCCACAGGATTTTGGCCAACAG 899
Qy	850	cttctcaactcttgatagatacaactccacatccacaggaagcctccacaggctgaatgcc 909
Db	898	CTTCTCAATCTTGAGATACAACTCCACCATCCAGAGAAAGCCTCCACAGGCTGAATGCC 839
Qy	910	ttccacgggttatctctccctccagtgctccactgcgagctgagggccatgacggcg 969
Db	838	TTCCACGGGTATATCTCCCTCAGTGTTCCTACTGCGGACTGAGGGCCATGACCGGCG 779
Qy	970	gatcgtgggagggcgctggcctcggttagcagtagcaagtggccttggaagtgtgactt 1029
Db	778	GATCGTGGGAGGGCGCTGGCGCTCGGATAGCAAGTGGCGCTTGGCAAGTGAGTCTGCATTT 719
Qy	1030	cggcaccacacacatctgtgagcagcagctcattgagccagtggtgctcactgcgc 1089
Db	718	CGGACCAACCCACATCTGTGAGGACGCTCATTTAGCGCCAGTGGGTGCTCACTGCCGC 659
Qy	1090	ccactgtctctctgacccgggagaggtctctgagggctggaaggtgtacgcggcgac 1149
Db	658	CCACTGTCTTCTGTGACCCGGAGAGGCTCTGGAGGCTGGAAGTGTACGGGGGCAC 599
Qy	1150	cagcaactctgacagttgctgaggcagcctccattgcgagatcatcatcaacagcaa 1209
Db	598	CAGCAACTCTGACCAAGTGTGCTGAGGACGCTTCCATTTGCCGAGATCATCATCAACAGCAA 539
Qy	1210	ttaccagtagagagacactatgacatgcctcctcctcgtcggctgtccagccctgac 1269
Db	538	TTACACCGATGAGGAGGACGACTATGATACATGCCCTCATGGGCTGTCCAAGCCCTTGAC 479
Qy	1270	cctgtccgctcacatccacccctgttgcctcccatgcatggacagaccttttagcctcaa 1329
Db	478	CCTGTCCGCTCACATCCACCTGTCTTGCCTCCCATGCATGGACAGACCTTTAGCCTCAA 419
Qy	1330	tgagacctgtggtatcacaggttttgcaagaccaggagagacagatgacaagacatcccc 1389
Db	418	TGAGACCTGTGTGATCATCAGGCTTTGGCAAGACCAGGAGACAGATGACAAGACATCCCC 359
Qy	1390	cttctccggaggtgaggtcaactcatctcactcactcaagaataatgactacttgg 1449
Db	358	CTTCTCCGGAGGTGAGGTCATCTCATCGACTTCAAGAAATGCAATGACTACTTTGGT 299
Qy	1450	ctatgacagttaccttaccctcccaagtagatgtgtgctggggaccttctgtggggcagaga 1509
Db	298	CTATGACAGTTACCTTACCCCAAGGATGTGTGTGCTGGGGACCTTCTGGGGGCACAGA 239
Qy	1510	ctcctgcagggagagacagcgggggctctgtctgtgagcagaaacacgcgtggtacct 1569
Db	238	CTCTCGCAGGGAGACAGCGGGGGCGCTCTTGTCTGTGTGAGCAGAACACCGCTGGTACCT 179
Qy	1570	ggcaggtgtaccagctgagggcagcagctgtggcagagaaacaaacctgtgtgtacac 1629
Db	178	GGCAGGTGTACACAGCTGGGGCAGACGCTGTGGCCAGAGAAACAAACCTTGTGTGTACAC 119
Qy	1630	caagtgacagaagtcttccctggatttacagcaagatggagagcaggtgtgcattca 1688
Db	118	CAAAAGTGACAGAAAGTTCTTCCCTGGATTTACAGCAAGATGGAGAGCGGATGCGATTCA 60

RESULT 7

AAD02322

ID AAD02322 standard; cDNA; 1222 BP.

XX AC AAD02322;

XX DT 28-MAR-2001 (first entry)

XX DE Human serine protease #3, encoding HATEE38 cDNA clone.

XX KW Human; serine protease; osteopathic; immunosuppressive; antiallergic;
antiinflammatory; cytostatic; cardiant; neuroprotective; nootropic;
neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
treatment; bone formation disorder; osteoporosis; arthritis; cancer;
connective tissue disorder; autoimmune disorder; wound healing; asthma;
systemic lupus erythematosus; male reproductive system disorder;
testicular cancer; digestion and food absorption disorder; arrhythmia;
Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
cardiovascular disorder; ocular disorder; drug screening; ss.

OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 76..1194

XX FT /*tag= a

XX FT /product= "Human serine protease #3 from clone HATEE38"

XX FT CDS 76..963

XX FT /*tag= b

XX FT /product= "Human serine protease #3 fragment from

XX FT clone HATEE38"

XX FT /note= "Does not include stop codon"

XX FT /partial

XX PN W0200068247-A2.

XX XX 16-NOV-2000.

XX PF 05-MAY-2000; 2000MO-US12207.

XX PR 07-MAY-1999; 99US-0133239.

XX PR 20-MAY-1999; 99US-0135163.

XX PR 03-AUG-1999; 99US-0147005.

XX PR 09-SEP-1999; 99US-0152935.

XX PR 01-NOV-1999; 99US-0162979.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Shi Y, Young PE, NI J;

XX XX WPI: 2000-679799/66.

XX DR P-PSDB; AAY72092, AAY72108.

XX PT New nucleic acid molecules encoding human serine protease polypeptides,
useful for diagnosis, prevention and/or treatment of disorders e.g.
osteoporosis, lupus erythematosus and Alzheimer's -

XX PS Claim 1; Page 271-272; 289pp; English.

XX CC The present cDNA sequence encodes human serine protease #3 from clone
HATEE38 (ATCC Deposit No: PTA27).

XX CC The invention relates to human serine proteases and their cDNA clones.
It is used in methods for the diagnosis, prevention and treatment of
various disorders related to serine protease such as bone formation
disorders (osteoporosis), connective tissue disorders (arthritis),
autoimmune disorders (systemic lupus erythematosus), wound healing, male
reproductive system disorders (testicular cancer), digestion and food
absorption disorders (Crohn's disease), neurodegenerative diseases
(Alzheimer's disease), behavioral disorders (Tourette's syndrome),
proliferative and cancerous conditions (acute myelogenous leukaemia),
allergic reactions (asthma), cardiovascular disorders (arrhythmia),

Qy 61 tggagcatctccagccaggcatctccagctgggacacctccagccggcgccgcatctccagc 120
|||||
Db 141 tggagcatctccagccaggcatctccagctgggacacctccagccggcgccgcatctccagc 200
|||||
Qy 121 ccaggcatctccagccaggcatctccagctgggacacctccagccggcgccgcatctccagc 180
|||||
Db 201 ccaggcatctccagccaggcatctccagctgggacacctccagccggcgccgcatctccagc 260
|||||
Qy 181 ccaggcatctccagctgtacacctccagccggcgccgcatctccagccggcgccgcatctccagc 240
|||||
Db 261 ccaggcatctccagctgtacacctccagccggcgccgcatctccagccggcgccgcatctccagc 320
|||||
Qy 241 ccaggcatctccagccc 257
|||||
Db 321 ccaggcatctccagccc 337
|||||

RESULT 9
AAS91044
ID AAS91044 standard; cDNA; 1212 BP.
XX AC
XX AAS91044;
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #26848.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX
XX 23-AUG-2000; 2000US-0649167.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR
XX
XX P-PSDB; ABG26857.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
PT
PT
PT
PT
PT
PS
XX
XX Claim 1; SEQ ID No 26848; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1212 BP; 275 A; 348 C; 328 G; 261 T; 0 other;

Query Match 14.3%; Score 250; DB 23; Length 1212;
Best Local Similarity 99.7%; Pred. No. 4.2e-98;
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 804 gtgctaccggacacccagggtgtccacagggattttgcaacagcttctcaatcttga 863
|||||

Db 404 gtgctaccggacacccagggtgtccacagggattttgcaacagcttctcaatcttga 463
|||||

Qy 864 gatacaactccaccatccaggaaagcctccacaggtgtgaatccctccagcggtata 923
|||||

Db 464 gatacaactccaccatccaggaaagcctccacaggtgtgaatccctccagcggtata 523
|||||

Qy 924 tctctctccagtgctccactggcgactggcgccatgacggcgcgatcgtggaggagg 983
|||||

Db 524 tctctctccagtgctccactggcgactggcgccatgacggcgcgatcgtggaggagg 583
|||||

Qy 984 cgtggcctcggatagcaagtgccttggcaagtgaatcgtgcactcggcacaccacaca 1043
|||||

Db 584 cgtggcctcggatagcaagtgccttggcaagtgaatcgtgcactcggcacaccacaca 643
|||||

Qy 1044 tctgtgagcagcactcatigaccccgatgggtgctcactcgcgccactgctcttcg 1103
|||||

Db 644 tctgtgagcagcactcatigaccccgatgggtgctcactgctcttcg 703
|||||

Qy 1104 t 1104

Db 704 t 704

RESULT 10

AAL02725

ID AAL02725 standard; DNA; 926 BP.

XX AC

XX AAL02725;

XX
XX 21-NOV-2001 (first entry)

DT
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 5413.

DE
XX
XX Human reproductive system related antigen; reproductive system disorder;

XX
XX
XX Human; cancer; gene therapy; ds.

KW
XX
XX Homo sapiens.

OS
XX
XX WO200155320-A2.

PN
XX
XX 02-AUG-2001.

PD
XX
XX 17-JAN-2001; 2001WO-US01339.

XX
XX
XX 31-JAN-2000; 2000US-0179065.

XX
XX
XX 04-FEB-2000; 2000US-0180628.

XX
XX
XX 24-FEB-2000; 2000US-0184564.

XX
XX
XX 02-MAR-2000; 2000US-0186350.

XX
XX
XX 16-MAR-2000; 2000US-0189874.

XX
XX
XX 17-MAR-2000; 2000US-0190076.

XX
XX
XX 18-APR-2000; 2000US-0198123.

XX
XX
XX 19-MAY-2000; 2000US-0205515.

XX
XX
XX 27-JUN-2000; 2000US-0209467.

XX
XX
XX 28-JUN-2000; 2000US-0214886.

XX
XX
XX 30-JUN-2000; 2000US-0215135.

XX
XX
XX 07-JUL-2000; 2000US-0216647.

XX
XX
XX 07-JUL-2000; 2000US-0216880.

XX
XX
XX 11-JUL-2000; 2000US-0217487.

XX
XX
XX 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen -
is used in preventing, treating or ameliorating a medical condition -

Disclosure; SEQ ID NO 5413; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders,
including cancer. The present sequence is a genomic sequence encoding a
protein of the invention.

Sequence 926 BP; 189 A; 225 C; 268 G; 244 T; 0 other;

Query Match

8.9%; Score 156; DB 22; Length 926;

Best Local Similarity 100.0%; Pred. No. 1.5e-57;

Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR	01-SEP-2000;	2000US-02293344;
PR	01-SEP-2000;	2000US-02293345;
PR	05-SEP-2000;	2000US-02295059;
PR	05-SEP-2000;	2000US-02295059;
PR	06-SEP-2000;	2000US-02295113;
PR	06-SEP-2000;	2000US-02295113;
PR	06-SEP-2000;	2000US-02304338;
PR	08-SEP-2000;	2000US-02312423;
PR	08-SEP-2000;	2000US-02312423;
PR	08-SEP-2000;	2000US-02312443;
PR	08-SEP-2000;	2000US-02312443;
PR	08-SEP-2000;	2000US-02314113;
PR	08-SEP-2000;	2000US-02314114;
PR	08-SEP-2000;	2000US-02320800;
PR	08-SEP-2000;	2000US-02320801;
PR	12-SEP-2000;	2000US-02319668;
PR	14-SEP-2000;	2000US-02323977;
PR	14-SEP-2000;	2000US-02323977;
PR	14-SEP-2000;	2000US-02323989;
PR	14-SEP-2000;	2000US-02323989;
PR	14-SEP-2000;	2000US-02324000;
PR	14-SEP-2000;	2000US-02324000;
PR	14-SEP-2000;	2000US-02324011;
PR	14-SEP-2000;	2000US-02324011;
PR	14-SEP-2000;	2000US-02324063;
PR	14-SEP-2000;	2000US-02324063;
PR	14-SEP-2000;	2000US-02324064;
PR	21-SEP-2000;	2000US-02342223;
PR	21-SEP-2000;	2000US-02342223;
PR	21-SEP-2000;	2000US-02342274;
PR	25-SEP-2000;	2000US-02349977;
PR	25-SEP-2000;	2000US-02349977;
PR	26-SEP-2000;	2000US-02349988;
PR	26-SEP-2000;	2000US-02349988;
PR	27-SEP-2000;	2000US-02358334;
PR	27-SEP-2000;	2000US-02358336;
PR	29-SEP-2000;	2000US-02363627;
PR	29-SEP-2000;	2000US-02363627;
PR	29-SEP-2000;	2000US-02363668;
PR	29-SEP-2000;	2000US-02363669;
PR	29-SEP-2000;	2000US-02363670;
PR	02-OCT-2000;	2000US-02368002;
PR	02-OCT-2000;	2000US-02370337;
PR	02-OCT-2000;	2000US-02370338;
PR	02-OCT-2000;	2000US-02370339;
PR	02-OCT-2000;	2000US-02370340;
PR	13-OCT-2000;	2000US-02399935;
PR	13-OCT-2000;	2000US-02399937;
PR	20-OCT-2000;	2000US-02409600;
PR	20-OCT-2000;	2000US-02412221;
PR	20-OCT-2000;	2000US-02412221;
PR	01-NOV-2000;	2000US-02411785;
PR	20-OCT-2000;	2000US-02411785;
PR	20-OCT-2000;	2000US-02417867;
PR	20-OCT-2000;	2000US-02417877;
PR	20-OCT-2000;	2000US-02418008;
PR	20-OCT-2000;	2000US-02418009;
PR	20-OCT-2000;	2000US-02418026;
PR	01-NOV-2000;	2000US-02446617;
PR	08-NOV-2000;	2000US-02464674;
PR	08-NOV-2000;	2000US-02464674;
PR	08-NOV-2000;	2000US-02464675;
PR	08-NOV-2000;	2000US-02465225;
PR	08-NOV-2000;	2000US-02465226;
PR	08-NOV-2000;	2000US-02465227;
PR	08-NOV-2000;	2000US-02465228;
PR	08-NOV-2000;	2000US-02464678;
PR	08-NOV-2000;	2000US-02464678;
PR	08-NOV-2000;	2000US-02465223;
PR	08-NOV-2000;	2000US-02465232;
PR	17-NOV-2000;	2000US-02492007;
PR	17-NOV-2000;	2000US-02492007;
PR	17-NOV-2000;	2000US-02492009;
PR	17-NOV-2000;	2000US-02492010;
PR	17-NOV-2000;	2000US-02492111;
PR	17-NOV-2000;	2000US-02492111;
PR	17-NOV-2000;	2000US-02492112;
PR	17-NOV-2000;	2000US-02492112;

PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249222.
PR 17-NOV-2000; 2000US-0249223.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249226.
PR 17-NOV-2000; 2000US-0249227.
PR 17-NOV-2000; 2000US-0249228.
PR 17-NOV-2000; 2000US-0249229.
PR 17-NOV-2000; 2000US-0249230.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX Disclosure; SEQ ID NO 5412; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 926 BP; 190 A; 224 C; 267 G; 245 T; 0 other;

Query Match 8.3%; Score 145; DB 22; Length 926;
Best Local Similarity 100.0%; Pred. No. 8.3e-53;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1376 gacaagacatccccctctccgaggaggtgcaggtcaatctcagcttcaagaaatgc 1435
Db 7 gacaagacatccccctctccgaggaggtgcaggtcaatctcagcttcaagaaatgc 66

Qy 1436 aatgactacttggtctatgacagttaacttaccacccaggtatgtgtgtgtggacatt 1495
Db 67 aatgactacttggtctatgacagttaacttaccacccaggtatgtgtgtgtggacatt 126

Qy 1496 cgtgggggagagactctcgcagg 1520
Db 127 cgtgggggagagactctcgcagg 151

RESULT 14
AAD05802/c
ID AAD05802 standard; DNA; 29 BP.
XX
AC AAD05802;
XX
DT 31-JUL-2001 (first entry)
XX
DE Human endothelialase 2 gene specific 5' RACE PCR primer.
XX

KW Human; endothelialase 2; protease domain; cytostatic; vulnery; wound;
KW notropic; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasm; ulcer; PCR primer; RACE;
KW rapid amplification of cDNA end; ss.
XX Homo sapiens.
XX
PN WO200136604-A2.
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US31803.
XX
XX 18-NOV-1999; 99US-0166391.
PR 22-SEP-2000; 2000US-0234840.
XX (CORV-) CORVAS INT INC.
XX Madison EL, Ong EO;
XX WPI; 2001-336001/35.
XX
XX New nucleic acid encoding a protein comprising endothelialase activity
PT useful in the prevention and treatment of e.g. vascular malformations,
PT cardiovascular disorders, and chronic inflammatory disease -
XX
XX Example 2; Page 117; 152pp; English.
XX
XX The present DNA sequence is human endothelialase 2 gene specific 5' RACE
CC (rapid amplification of cDNA ends) PCR primer which is used for obtaining
CC a full-length cDNA clone encoding endothelialase 2.
CC The invention relates to an endothelialase protein, endothelialase protease
CC domain and their corresponding nucleic acid molecules. An endothelialase
CC protein or protease domain of it is useful for the treatment and
CC diagnosis of disorders associated with aberrant angiogenesis or undesired
CC neovascularisation. The undesired angiogenesis is associated with
CC disorders selected from solid neoplasm, vascular malformations and
CC cardiovascular disorders such as angiofibroma, angiolipoma,
CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
CC malformations, haemangiomas and vascular adhesions, dyschondroplasia
CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
CC periodontitis, psoriasis, hypertrophic scars, liver cirrhosis,
CC granulations-burns, rheumatoid arthritis, venous stasis ulcers,
CC osteoradionecrosis, postoperative wound repairs, circulatory disorders
CC systemic sclerosis and aberrant wound repairs, pyogenic granuloma and
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
CC dysmotility, sclerodactyly and teangiectasis, dermatological disorders
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
CC caused by ocular neovascular disease, corneal graft neovascularisation,
CC macular degeneration, retinopathy of prematurity, retrolental
CC fibroplasia and corneal neovascularisation. The nucleic acids of the
CC invention are also used in gene therapy. The invention also provides
CC method for screening compounds that modulate angiogenesis.
XX
SQ Sequence 29 BP; 7 A; 4 C; 15 G; 3 T; 0 other;

Query Match 1.7%; Score 29; DB 22; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1273 gtccgtccacatccaccctgtctgctcc 1301

Db 29 GTCCGCTCACATCCACCTGCTGCCCTCC 1
|||||

RESULT 15

AAD05809/c

ID AAD05809 standard; DNA; 29 BP.

XX AC AAD05809;

XX DT 31-JUL-2001 (first entry)

XX DE Human endothelias 2-specific anti-sense RT-PCR primer.

XX

KW Human; endothelias 2; protease domain; cytosolic; vulnary; wound;
KW neotrophic; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disease; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
KW liver cirrhosis; osteoradionecrosis; systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasia; ulcer; RT-PCR primer; ss.
XX Homo sapiens.

OS

XX WO200136604-A2.

XX PD 25-MAY-2001.

XX PF 17-NOV-2000; 2000WO-US31803.

XX PR 18-NOV-1999; 99US-0166391.

XX PR 22-SEP-2000; 2000US-0234840.

XX PA (CORV-) CORVAS INT INC.

XX PI Madison EL, Ong EO;

XX DR WPI; 2001-336001/35.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
CC caused by ocular neovascular disease, corneal graft neovascularisation,
CC macular degeneration, retinopathy of prematurity, retrolental
CC fibroplasia and corneal neovascularisation. The nucleic acids of the
CC invention are also used in gene therapy. The invention also provides
CC method for screening compounds that modulate angiogenesis.
XX
SQ Sequence 29 BP; 7 A; 4 C; 15 G; 3 T; 0 other;

Query Match 1.7%; Score 29; DB 22; Length 29;

Best Local Similarity 100.0%; Pred. No. 0.013;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1273 gtccgctcacatccacccctgctgctcc 1301

DB 29 GTCCGCTCACATCCACCTGCTGCCCTCC 1

Search completed: August 19, 2002, 23:53:10

Job time: 8967 sec

New nucleic acid encoding a protein comprising endothelias activity
useful in the prevention and treatment of e.g. vascular malformations,
cardiovascular disorders, and chronic inflammatory disease -

Example 2; Page 120; 152pp; English.

The present DNA sequence is human endothelias 2-specific anti-sense RT
(Reverse Transcriptase)-PCR primer which is used for amplifying a 422-bp
fragment from the single-stranded cDNAs of human umbilical vein
endothelial cells (HUVEC). This 422-bp fragment spans the scavenger
receptor cysteine-rich domain and the serine protease domain.
The invention relates to an endothelias protein, endothelias protease
domain and their corresponding nucleic acid molecules. An endothelias
protein or protease domain of it is useful for the treatment and
diagnosis of disorders associated with aberrant angiogenesis or undesired
neovascularisation. The undesired angiogenesis is associated with
disorders selected from solid neoplasia, vascular malformations and
cardiovascular disorders such as angiofibroma, angiolipoma,
atherosclerosis, restenosis/reperfusion injury, arteriovenous
malformations, haemangiomas and vascular adhesions, dyschondroplasia
with vascular hamartomas (Faucci's syndrome), hereditary haemorrhagic
telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
syndrome, chronic inflammatory diseases such as diabetes mellitus,
haemophilic joints, inflammatory bowel disease, nonhealing fractures,
periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
granulations-burns, hypertrophic scars, liver cirrhosis,
osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
systemic sclerosis and aberrant wound repairs, circulatory disorders
Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
dyomeitoly, sclerodactyly and teangiectasis, dermatological disorders
such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 20:52:13 ; Search time 3210.3 Seconds
(without alignments)
11394.445 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctcagagaccatggagaggg.....ggctgtgtgactcgagaaa 1748

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description

1	1678	96.0	1689	6	AX149579	Sequence
2	1663	95.1	2067	6	AX149581	Sequence
3	1621	92.7	2393	9	AB048796	Homo sapi
4	1253	71.7	1314	6	AX190543	Sequence
5	1253	71.7	1314	6	AX190546	Sequence
6	1188	68.0	3324	9	AB048797	Homo sapi
7	1049	60.0	2192	9	AK027798	Homo sapi
8	1019	58.3	1078	6	AX190545	Sequence
9	227	13.0	160222	9	AP002962	Homo sapi
10	227	13.0	184926	9	AP000757	Homo sapi
11	156	8.9	111094	2	AP000683	Homo sapi
12	152	8.7	450	6	AX246661	Sequence
13	49	2.8	2125	10	BC010843	Mus muscu
14	29	1.7	29	6	AX149587	Sequence
15	29	1.7	29	6	AX149595	Sequence
16	27	1.5	27	6	AX149589	Sequence
17	27	1.5	27	6	AX149591	Sequence
18	26	1.5	26	6	AX149588	Sequence
19	25	1.4	25	6	AX149596	Sequence
20	25	1.4	36	6	AX149592	Sequence
21	24	1.4	39	6	AX149601	Sequence
22	24	1.4	969	9	HUMSERPROT	L33404 Human strat
23	24	1.4	986	6	A42048	Sequence 1
24	24	1.4	986	6	AR053966	Sequence
25	24	1.4	986	6	AR084988	Sequence
26	24	1.4	1089	6	AR103963	Sequence
27	24	1.4	1089	6	AR103964	Sequence
28	24	1.4	9729	9	AF332583	Homo sapi
29	24	1.4	9909	9	AF166330	Homo sapi
30	24	1.4	109124	2	AC011483	Homo sapi
31	24	1.4	217346	2	AC027602	Homo sapi
32	24	1.4	230000	9	AF243527	Homo sapi
33	23	1.3	23	6	AX190548	Sequence
34	23	1.3	164780	2	AC109583	Homo sapi
35	23	1.3	179240	2	AC094602	Rattus no
36	22	1.3	22	6	AX149594	Sequence
37	22	1.3	2009	3	AY070572	Drosophil
38	22	1.3	81092	2	AC014998	Drosophil
39	22	1.3	160005	3	AC007176	Drosophil
40	22	1.3	168909	2	AC017470	Drosophil
41	22	1.3	169289	3	AC007145	Drosophil
42	22	1.3	265605	3	AE003662	Drosophil
43	22	1.3	299620	3	AE003429	Drosophil
44	21	1.2	21	6	AX190547	Sequence
45	21	1.2	21	6	AX190549	Sequence

ALIGNMENTS

RESULT	1	AX149579	Sequence 3 from Patent WO0136604.	1689 bp	DNA	linear	PAT 08-JUN-2001
LOCUS	AX149579	Sequence 3 from Patent WO0136604.					
DEFINITION	AX149579	Sequence 3 from Patent WO0136604.					
ACCESSION	AX149579	Sequence 3 from Patent WO0136604.					
VERSION	AX149579.1	GI:14348013					
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1689)
Madison, E.L. and Ong, E.O.
Nucleic acids encoding endotheliases, endotheliases and uses thereof

Patent: WO 0136604-A 3 25-MAY-2001;
CORVAS INTERNATIONAL, INC. (US)
Location/Qualifiers

1. .1689
/organism="Homo sapiens"
/db_xref="taxon:9606"

1. .1689
/note="DNA encoding a transmembrane serine protease"

(Endotheliase 2-S) protein"

```
/codon_start=1
```

```
/protein_id="CAC41220.1"
```

```
/db_xref="GI:14348014"
```

/translation="MERDSHGNA

SPAGTTPGRASPAQASVAGTPGPRASPCGRASPAQASPARASPAISLSPSSCSGRSSLS
 RASVSTSPRYLYRNPATVPGRVPSRSPARSAPATRATRSPGTSPLKPTWREGQKQ
 LPLIGCVLLIALVSLILPQFWGHTGRIKQGRSCPHAVRQGVQDCKSKDSE
 LGCVREDQSLKLYIGSGSHWLPICSSNNNDVSBKTCQQLGFSAHRTTVEIARD
 FANFSLRYNSITQESLRSHSPSRYLSQSHCGRTMGRTGVGLASQSKWEP
 QVSLHFTTICGGTLDIAOWMLTAACHFVTRFKVLEKMGHTVAGTSNLHQLPEAS1
 AELIINSNTDEEDYDALMKPLFLSAH1HPACLPKHGQVTSNLHQLPTGFGK
 TRPTEQDTSFPLREQVNLDFKCNLDVYDSTVYTRMWCAGDLRGDRSCQSDGG
 PLVCEONNRYLAGVTSWGTGCGORNGKGVYTKTVLEVPLNYSKMESEBRYFKS

529 c 470 a 323 f

ORIGIN

Query Match		96.0%;	Score 1678;	DB 6;	Length 1689;
Best Local Similarity		100.0%;	Prod. Mismatch No. 0;		
Matches 1678;		Conservative	0;	Mismatches	0;
QY	11	atgagagggacagccacgaggaatgcattctccagacaagaacaccttcagctggagcatct	70		
DB					
DB	1	ATGGAGAGGACACGCCACGGGAATGCATCTCCAGCAAGAACACCTTTCAGCTTGGAGCATCT	60		
QY	71	ccagccaggcatctccagctgggacacctccagcccgggcgatctccagcccgagcatct	130		
DB	61	CCAGCCAGGCATCTCCAGTGGGACACCTCCAGCCGGGCATCTCCAGCCAGGCATCT	120		
QY	131	ccagccaggcatctccagctgggacacctccgggcccggcgcatctccagcccgagcatct	190		
DB	121	CCAGCCAGGCATCTCCAGTGGGACACCTCCGGCCGGGCATCTCCAGCCAGGCATCT	180		
QY	191	ccagctggtacacctccaggccgggcatctccagcccgggcgatctccagcccgagcatct	250		
DB	181	CCAGCTGGTACACTCCAGGCCGGGCATCTCCAGCCGGGCATCTCCAGCCAGGCATCT	240		
QY	251	ccagccgggcatctcggctctggcatcatttccagggtctcctcatccgggaggtcatca	310		
DB	241	CCAGCCCGGGCATCTCCGGCTTCGGCATCACTTTCAGGTCCTCATCCGGCAGGTCATCA	300		
QY	311	tcgcgcaggttcagctcgggtgacacctccccaccagagtgtaacctgtttagagcaaca	370		
DB	301	TCGCGCAGGTCAGCTCGGTGACAAACCTCCCCACACAGAGTGTACTTGTGTAGAGCAACA	360		
QY	371	ccagtgggggctgtaccatctcgatcatctctgccagggtcagcaccagcaacccagggcc	430		
DB	361	CCAGTGGGGCTGTATCCATCCGATCATCTCTGCCAGTCAGCACCAGCAACCCAGGGCC	420		
QY	431	acagggagagccaggtagcagcctccaaagtctcacttgcggggagggccagaagcag	490		
DB	421	ACCAGGGAGAGCCAGGATACGCGCTCGCAAGTTCACCTTGGCGGGAGGGCCAGAAGCAG	480		
QY	491	ctaccgctcatcgggtgcgtgctcctcctcattgccctgtggtttcgctcatcatctc	550		
DB	481	CTACCGGCTCATGCGGTGCGTGTCTCTCTCTCATGCCCCCTGTGGTTCCTCATCATCTCTC	540		
QY	551	ttccagttctggcaggccacacagggatcagggtacagagcgagggagagctgtccc	610		
DB	541	TTCCAGTCTTGGCAGGGCCACACAGGGCATCAGGTACAGGAGCAGAGGGAGCTGTCCC	600		
QY	611	aagcagccttctcgctgtgacgggtgtgtgagctgcaagctgaagctgaagctgagcagctgggc	670		
DB	601	AAGCACGCTGTTCGCTGTGACGGGGTGGTGGACTTGC AAGCTGAAGTGAAGTGAAGCTGGGC	660		
QY	671	tgcgtgaggttttactggtggacaagctctcgtcttaaaatctactctgggtcctcccatcag	730		
DB	661	TGCGTGAGGTTTGACTGGGACAGTCTCTGCTTAAATCTACTCTGGGTCTCTCCCATCAG	720		
QY	731	tgccttcccatctgtagcagcaactggaaatgactcttactcagagaagacgtgccagcag	790		
DB	721	TGCGTTCCCATCTGTAGCAGCAACTTGAATGACTCTTACTTACAGAGAAGACTGCCAGCAG	780		

REFERENCE 1 (bases 1 to 2067)
 AUTHORS Madison E.L. and Ong, E.O.
 TITLE Nucleic acids encoding endotheliases, endotheliases and uses thereof
 JOURNAL Patent: WO 0136604-A 5 25-MAY-2001;
 CORVAS INTERNATIONAL, INC. (US)
 FEATURES Location/Qualifiers
 source 1..2067
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 CDS 1..2067
 /note="DNA sequence encoding a transmembrane serine protease (endothelase-L) protein"
 /codon_start=1
 /protein_id="CAC41221.1"
 /db_xref="GI:14348016"
 /translation="MERDISHGNASPARTPSAGASPAQASPAQTTPGGRASPAQASPAQA
 SPAGTPPGRASPAQASPAQTTPGGRASPAQASPARASPAQASLSRSSGSRSSA
 RLSAVTSPTRVLYLKVATPGVIRSPARSAPATRPSPKVRGTRSLPKFTWREGQKQ
 LPLIGVLLI IALVGLII ILFPQWQGTGIRYKEQRESCEPKHACRQDGVVDCRKLKSGE
 LGCVRWDKSLKIIYSGSSHWQPLPICSSNWDNSYSEKTCQOLCFESAHTTVEAHRD
 FANSFILIYNTSIQESLHRSECPORSYISLQCSHGCLRAMTGRI VGGALASDKWPI
 QVSLHFGTHICGGTILDAQWLTAACHFEVTREREKYLEGWKYAGTSNLHQLPEASII
 AEIIINSNTDESDDI IALMLSKPLTISAHILPACLPMHGGTFSNTECWTIGTGGK
 TRETFDTSPLREDDVNDILFKKNDIYLDVSYLTPRMCAQGLDRGDRSDCGSGG
 PLVCEQNNMYLAGVTSWGTGQQRNKPQVYTKVTELPWIKYSKMENRAQVRVKATY
 RPRQQLLGRCSPIRSLFKVAMDENFVSVSAEDFVIFVFIKHLCMGIRSRSWPPPALFY
 LVFLIFLLILLSFLKNTSDSILTLTFTAVTRMLPENYHSRPFPELKFHQNLTLTIK
 EYKVIK"
 BASE COUNT 463 a 604 c 540 g 460 t
 ORIGIN
 Query Match 95.1%; Score 1663; DB 6; Length 2067;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 atggagaggacacccaggggaatgcattctccagcaagaacacaccttcagctggagcatct 70
 |||||
 Db 1 ATGGAGAGGGACAGCCACGGGAATGCATCTCCAGCAAGAACACCTTCAGCTGGAGCATCT 60
 QY 71 ccagccaggcatctcagctgggacaacctccaggccgggcattctccagccaggcatct 130
 |||||
 Db 61 CCAGCCAGGCATCTCCAGCTGGGACACCTCCAGCCGGGCATCTCCAGCCAGGCATCT 120
 QY 131 ccagccaggcatctcagctgggacaacctccgggcccggcattctccagccaggcatct 190
 |||||
 Db 121 CCAGCCAGGCATCTCCAGCTGGGACACCTCCGGGCCGGGCATCTCCAGCCAGGCATCT 180
 QY 191 ccagctgtgtacacctccagccggggcatctctccaggccgggcattctccagccaggcatct 250
 |||||
 Db 181 CCAGCTGTACCTCCAGCCGGGCATCTCCAGCCGGGCATCTCCAGCCAGGCATCT 240
 QY 251 ccagccgggcattctcgggtcttggaatacttccagggtctcattccagccaggcatca 310
 |||||
 Db 241 CCAGCCGGGCATCTCCGGGTCTGGCATCATCTTCCAGGTCTCTATCCGGCAGGTCTCA 300
 QY 311 tcgcagggtcagcctcggtgacaacctcccccaacagagtacattgttagagaaca 370
 |||||
 Db 301 TCGCCAGGTGAGCTCGGTGACAACTCCCCAAACAGAGTGTACCTTGTTAGAGCAACA 360
 QY 371 ccagtgggggctgtaccacatccgatcatctcctgccagggtcagcaccagcaaacagggcc 430
 |||||
 Db 361 CCAGTGGGGCTGTATCCATCCGATCATCTCCTGCCAGGTTCAGCACCAGCAACAGGCC 420
 QY 431 accagggagagccaggatcagagcctgcccaagttcacctggcgggagggccagaagcag 490
 |||||
 Db 421 ACCAGGAGAGCCAGGTAGAGGCTGCCCAAGTTCACCTGGCGGAGGGGCCAAGACGAG 480
 QY 491 ctaccgctcatcgggtgcgtgcctcctcatctcagcctggtggtttgcctcatcctc 550
 |||||
 Db 481 CTACCGCTCATCGGGTGCgtgcctcctcctcctcctcctcctcctcctcctcctcctcctc 540

QY	551	ttccagttcttgcgagggccacacagggatacaggtacaaggagcagaggagagctgtccc	610
DB	541	TTTCCAGTTCTGCGAGGGCCACACAGGGATCAGGTACAAGGAGCAGAGGGAGAGCTGTCCC	600
QY	611	aagcaagctgttcgctgtgacaggggtggtggactcaagctgaagagtgaagagctgggc	670
DB	601	AAGCAGCGTTCGTGTGACGGGGTGGTGACGTGCAAGCTGAAGAGTGAAGAGCTGGGC	660
QY	671	tgcgtgaggtttgactgggacaagtctctgtctataaatactactctgggttctctccactcag	730
DB	661	TGCGTGGAGTTTACTTGGGACAAAGTCTCTGTCTTAAATCTACTCTGGTCTCTCCATCAG	720
QY	731	tggcttccccatctgtagcagcaactggaatgactctactctacagagaagacctgccaagcag	790
DB	721	TGGCTTCCATCTGTAGCAGCAACTGGAATGACTCTCTACTCAGAGAAGACCTGCCAGCAG	780
QY	791	ctgggtttcgaagtgctcacaggacaaccgaggtttgccacagggatittgccaaacagc	850
DB	781	CTGGGTTTCGAGAGTGCTCACCGGACAACCGAGGTTGCCACAGGGATTTGGCAACAGC	840
QY	851	ttctcaattcttgagatacaactccaccatccaggaagcctccacaggtctgtaatgccct	910
DB	841	TTCTCAATCTTGAGATACAACCTCCACCATCCAGGAAGCCTCCACAGGTCGTGAATGCCCT	900
QY	911	tccagcggtatatactccctccagtgttcccaactgcggaactgagggccatgacctggcgcg	970
DB	901	TCCACAGCGTATATCTCCTCCAGTGTTCCTCCTGCGACTGAGGGCCATGACCGGGCGG	960
QY	971	atcgtggagggcgctggcctcgaatacgaagtggccttgccaagtgaagtgaactgcacttc	1030
DB	961	ATCTGTGGAGGGCGCTGCGCTCGGATGACAGTGGCTTGGCAAGTGAAGTCTGCACTTC	1020
QY	1031	ggcacaccacatatctgtgagggcacgctcattgacgcccagtggggtgctcaactgcgcgc	1090
DB	1021	GGCACCCACATCTGTGGAGCAGCTCATTTGACGCCCGAGTGGGTGCTCACTGCGGCC	1080
QY	1091	caactgtcttctgtgacccgggagaaggttcctggaggggtgaaagdtgtacgcgggcaac	1150
DB	1081	CACCTGCTTCTTCGTACCCGGGAGAGGTCTGGAGGGCTGGAAGGTGTACGCGGGCAC	1140
QY	1151	aqcaacctgcacagttgcctgagcagcagcctccattgccagatcatcatcaacagcaat	1210
DB	1141	AGCAACCTGCAACAGTTGCTTGAGGAGCGCTCCATTTGCCGAGATCATCATCAACAGCAAT	1200
QY	1211	tacacgatgagaggagcaactatgacatcgccctcatcggtctgccaaagcccttgacc	1270
DB	1201	TACACGGATGAGGAGCAGCTATGACATCGCCCTCATCGGGCTGTCCAAGCCCCCTGACC	1260
QY	1271	ctgtccgtctacatccacccctgttgcctccccatgcatggaacagacctttagctccaat	1330
DB	1261	CTGTCCGCTACATCCACCTGCTTGCTGCCCTCCCATGCAATGGACAGACCTTTAGCTTCAAT	1320
QY	1331	gagaacctgtgatcacagcttttggcaagaccaggagagacagatgacaagacatcccc	1390
DB	1321	GAGACCTGCTGGATCACAGCTTTTGGCAAGACAGGAGACAGATCACAAGACATCCCC	1380
QY	1391	ttcctccgggaggtgcaggtcaactctcatcgacttcaagaaatgcaatgactacttggtc	1450
DB	1381	TTCTCTCGGGAGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTTACTTGGTC	1440
QY	1451	tatgacagttaccttaccocaaaggtatgattgtgtgtgggaacctcgtagggggcagagac	1510
DB	1441	TATGACAGTTACTTTACCCAAAGGATGATGTGTGTGGGGACCTTCGTGGGGCAGAGAC	1500
QY	1511	tectgcagggagacagcgggggaccttctgtgtgagcagaacaacgcgttgtagctcg	1570
DB	1501	TCCTGCCAGGGAGACAGCGGGGGGCTCTTTGTCTGTGACAAACAACCGCTGGTACCTG	1560
QY	1571	gcaggtgtcaccagctggggcacagctgtgcccagagaaacaaacctggtgtgtacacc	1630
DB	1561	GCAGGTGTACCAGCTTGGGGCACAGGCTGTGGCCAGAGAAACAAACCTGGTGTGTACACC	1620
QY	1631	aaagtgcagaagattcttccctggattttacagcaaatggaga	1673

Db 1621 AAGTGCACGAAGTTCTTCCTCGGATTATACAGCAAGATGGAGA 1663
|||||
RESULT 3
AB048796 2393 bp mRNA linear PRI 22-MAR-2001
LOCUS Homo sapiens mspl mRNA for membrane-type mosaic serine protease,
DEFINITION complete cds.
AB048796
AB048796.1 GI:13429969
Homo sapiens lung cDNA to mRNA, clone_lib:Marathon Ready cDNA
(Clontech).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kim,D.R., Sharmin,S., Inoue,M. and Kido,H.
TITLE Cloning and expression of novel mosaic serine proteases with and
without a transmembrane domain from human lung(1)
JOURNAL Blochim. Biophys. Acta 1518 (1-2), 204-209 (2001)
MEDLINE 21167393
REFERENCE 2 (bases 1 to 2393)
AUTHORS Kim,D.R., Inoue,M. and Kido,H.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2000) Hiroshi Kido, The University of Tokushima,
Division of Enzyme Chemistry, Institute for Enzyme Research;
Kuramoto-cho, Tokushima 770, Japan
{E-mail:kido@er.tokushima-u.ac.jp, Tel:81-88-633-7424,
Fax:81-88-633-7425}
FEATURES
source Location/Qualifiers
1..2393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="lung"
/clone_lib="Marathon Ready cDNA (Clontech)"
1..87
88..1833
/gene="mspl"
88..1833
/gene="mspl"
/codon_start=1
/transl_except=(pos:1573..1575,aa:His)
/transl_except=(pos:862..864,aa:Arg)
/transl_except=(pos:979..981,aa:His)
/product="membrane-type mosaic serine protease"
/protein_id="BAB39741.1"
/db_xref="GI:13429970"
/translation="MERDGHNSPARTSPASQAPASPAQSPAGTPPPGRASPAAQSPAQA
SPAGTPPGRASPAPAGTPPCRASPGPASPAPASPARASPALASLSRSSGSSEA
RSNVTSSTPRVLYRAPVPVAVPIRSSPARSAPATRAITRESPGTSLEKFTWRECKQ
LPLGVLLDLIALVSLILFLFWOGTHIRKEQRESCPRHVRCDGDVDCCLKSDE
LCRVDFDDKLKYISGSHOWLPICSSNWDYSSEKTROLGFESAHRTTEVAHRD
FANSFILRYNTIOESLRHSRCSQRVYISQCSCGLRAMTGRIVGALADSXKPW
QVSLHFHTGICGGTLIDAQWLTAARCFVTREKVLEGWKYAGTSNLHOLPEAASI
AEIINSNTDEEDDYDALMLRSLPLTSAIHHPACLPMHGOTFSLNETCWITGFG
TRETDDKTSFPIREVQNLIIDFKNCNDYLIVDSYLTIPRMKAGDLHGRDRSCQDGSG
PLVCQNNRWYLAGVTSMTGCGQRNKPFVITVKYTEVLPIWYSKMEVRSLOQDTAPS
LGTSSGDDPGGAPRV"
3'UTR 1834..2393
polyA_signal 2343..2348
BASE COUNT 553 a 692 c 627 g 521 t
ORIGIN

Query Match 92.7%; Score 1621; DB 9; Length 2393;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ctgagacattgaggagggagaccacgcgggaatgatcttcacagcaagaaccttcacgc
db 78 CTCAGAGACCATTGGAGGAGACAGCCAGGGAATGATCTCCAGCAAGAACACTTCAGC


```
QY 1141 cggggcaccagaaactgcacagtgctgctgagggcagcctccattgcccagatcatcat 1200
Db 1218 CGGGGACCAGCAACCTGCACAGTTGCTTGAGGCAGCCTCCATTGCCGAGATCATCAT 1277
QY 1201 caacagcaattacacccgatgaagagacacatgatgacatcgccctcatcgcgctgccaa 1260
Db 1278 CAACAGCAATTAACCGATGAGAGGACACATGATGACATGCGCCTCATCGCGTGTCCA 1337
QY 1261 gcccctgacccctgcgcgtcacatccacccctgctgctcccccacatgagacagacct 1320
Db 1338 GCCCTGACCCCTGTCGCTTCATCCACCTGCTTGCCTCCCATGATGAGACACCTT 1397
QY 1321 tagctcaatgagacctgctggtacacagccttggtggaagacagagagacagatgacaa 1380
Db 1398 TAGCCTCAATGAGACCTGCTGGATCACAGCTTTGGCAAGACAGCAGGAGACAGATGACAA 1457
QY 1381 gacatccccctccctccgggaggtgcaggtcaatctcatcgacttcaagaataatgaatga 1440
Db 1458 GACATCCCCCTTCCCTCCGGGAGGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGA 1517
QY 1441 ctacttggtctatgacaggttacttaccaccaaggtatgctgctgggagaccttcgtgg 1500
Db 1518 CTACTTGGTCTATGACAGTTACTTACCCCAAGGATGATGCTGCTGGGAGACCTTCTGG 1577
QY 1501 gggcagagactcctgccagggagacagcggggggcctcttctgtgagcagaacaaccg 1560
Db 1578 GGGCAGAGACTCCTGCCAGGGAGACAGCGGGGGGCTCTTGTGTGAGCAACAACCG 1637
QY 1561 ctggatcctggcaggtgtcaccagctgggcagagcgtggtgcccagagaaacacctgg 1620
Db 1638 CTGGTACCTTGGCAGGTGTCAACAGCTGGGGCACAGCTGTGGCCAGAGAAACAACCTGG 1697
QY 1621 tgtgtacacccaagtgcagaagttcttccctggatttaccagcaagatggag 1672
Db 1698 TGTGTACACCAAAAGTGACAGAAGTTCTTCCCTGGATTACAGCAAGATGGAG 1749

RESULT 4
AX190543
LOCUS AX190543 1314 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 14 from Patent WO0136645.
ACCESSION AX190543
VERSION AX190543.1 GI:15143854
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Quinn,K.E., Spytek,K.A., Majumder,K., Vernet,C., Burgess,C.,
Fernandes,E., Taupier,R.J., Rastelli,L. and Herrmann,J.L.
Serine/threonine kinase and serine protease polypeptides and
nucleic acids encoding same
JOURNAL Patent: WO 0136645-A 14 25-MAY-2001;
Curagen Corporation (US)
FEATURES
source
1..1314
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS
1..1266
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC50789.1"
/db_xref="GI:15143855"
/translation="MESPTSLPKFTWREGOKPLIGCVLLILVSLITLFOFWQ
GHTGIRYKEQRESCPKHVRCDGVVDCKLSDBLGCVRFDWDKSLKKIYSSSSHOWLP
ICSSNNDQISSEKTCQQLGFESAHRTEVAHRDFANSFSLIRNISTIOESLHRSCEPS
QRYLSLQCSHGLAMTRIGVGLAASDKSPWQVSLFHGTTHDGLDIAQWVITA
AHCFEYTRKLEGWKVYAGTSLNLQLPEAASIAEIIINSNYTDEEDDYDIAMLRSK
PLTSAHTHPACLPMHGOTFSLNETCWTITGFKTRTDKTSPLFREOVNLIDPKK
NDLYVDSYLTPRMWCAGDLRGGRDSGGPLVCEQNNRWYLAGVTSWGTGCGQR
NKQVYVTKVTLVLPWLYSKMESEVRFYKS"
BASE COUNT 288 a 384 c 376 g 266 t
```

```
ORIGIN
Query Match 71.7%; Score 1253; DB 6; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 436 ggagagccaggttacgagctgcccgaagttcacctggcgggagggccagaagcagctaac 495
Db 3 GGAGAGCCAGGTACGAGCCTGCCAAGTTCACCTGGCGGAGGGCCAGAGCAGCTACC 62
QY 496 gctcattgggtgctgctcctcctcattggcctggtggttctcctcctcctcctcctc 555
Db 63 GCTCATCGGTGGCTGCTCTCTCTCATTTGCCCTGGTGGTTTCGCTCATCTCTCTTCCA 122
QY 556 gttctgcagggccacacagggatcagggtacaaaggagcagagagagagctgtcccaaga 615
Db 123 GTTCTGGCAGGGCCACACAGGGATCAGGTACAAAGAGCAGAGGAGCTGTCCCAAGCA 182
QY 616 cgcgtgtcgtgtgacggggtggtgactgcaagctgaagagtgaagagctgggctgcgt 675
Db 183 CGCTGTTGCTGTGACGGGGTGGTGACTGCAAGCTGAAGAGTGAAGAGTGGGCTGGCT 242
QY 676 gagggttgactgggacaagtctctgtcttaaaatctactctggtcctccatcagtggt 735
Db 243 GAGGTTTGACTGGGACAAGTCTCTGCTTAAATCTACTCTGGTCTCTCCATCAGTGGCT 302
QY 736 tcccatctatgacagcaactggaatgactcctactcagagaagacactgccagcagctgg 795
Db 303 TCCCATCTGTAGCAGCACTGGAAATGACTCTACTCAGAGAAGACCTGCCAGCAGCTGG 362
QY 796 ttctgagagtgtcacccgacacacaggaggttgcaccaaggatttgcacaacagctctc 855
Db 363 TTTTCGAGAGTGTCTACCGGACAACCGAGGTTGCCACAGGGATTTTGCACAACAGCTTC 422
QY 856 aatcttgagatacaactccacatccacatccaggaaagcctccacaggtctgaatgccctc 915
Db 423 AATCTTGGATACAACATCCACCATCCAGGAAAGCCCTCCACAGCTCTGAATGCCCTTCCA 482
QY 916 gcggtatatctccctccagtggtccacactgcgagactgagggccatgacgggagctcgt 975
Db 483 GCGGTATATCTCCCTCCAGTGTTCCTCCACATGCGGACTGAGGGCCATGACGGGGGATCGT 542
QY 976 gggagggcgctggcctcgatagcaagtggccttgccaagtgtgactgtgcacttcggcac 1035
Db 543 GGGAGGGCGCTGGCCTCGATAGCAAGTGGCCTTGGCAAGTGCAGTCTGCACFTCGGCAC 602
QY 1036 caccacactctgtgagggcacgctcattgacgcccagtggtgctcactgcgcgcccactg 1095
Db 603 CACCCACATCTGTGGAGGCACGCTCATTCACGCCAGTGGGTGCTCCTGCGGCCCACTG 662
QY 1096 ctcttcgtgacccgggagaaggtcctgagggctggaaggtgtacgcgggcaaccagcaa 1155
Db 663 CTCTCTGCTGACCCGGGAGAAAGGTCTCTGGAGGGCTGGAAGGTGTACGGGGCACACGCAA 722
QY 1156 cctgcaccagtgtcgtgagcagcctccattgcccagatcatcatcaacagcaattacac 1215
Db 723 CCTGCACCAAGTTGCCCTGAGCAGACCTCCATTTGCCGAGATCATCATCAACAGCAATTA 782
QY 1216 cgatgagggagcactatgacatcgccctcatcgggctgtccaaagccctgacccctgtc 1275
Db 783 CGATGAGGAGGACGACTATGACATCGCCCTCATTCGGGCTGTCCAAAGCCCTTGACCTGT 842
QY 1276 cgtcacatccacctgctgtctccctccatgcatggacagacctttagctcctaatgagac 1335
Db 843 CGCTACATCCACCTCTGCTTCCCTCCCATGATGAGCAGACCTTTAGCCTCAATGAGAC 902
QY 1336 ctgctggatcacaggtcttggcaagaccaggagacagatgacaagacatcccccttct 1395
Db 903 CTGCTGGATCACAGGCTTTTGGCAAGACCAGGAGACAGATGACAAGACATCCCCCTTCT 962
QY 1396 ccgggaggtgcagggtcaattctcatcgacttcaagaataatgcaattgactgtctatga 1455
```

Db 963 CCGGAGGTGCGAGTCAATCTCATGCACTTCAAGAAATGCAATGACTACTTGGTCTATGA 1022
QY 1456 cagttacctaccaccaagatgatgtgtctgggggacctctctgtggggcagagactccctg 1515
|||||
Db 1023 CAGTTACCTTACCCCAAGATGATGTGCTGGGACCTTCTGGGGGCGAGAGACTCCCTG 1082
QY 1516 ccaggagacagcgggggcccctcttctgtgagcagaacaacccgctggttacctggcagg 1575
|||||
Db 1083 CCAGGAGACAGCGGGGGCCCTCTTCTGTGTGAGCAGAACACCCGCTGGTACCTGCGCAG 1142
QY 1576 tgtcaccaagctggggcacaaggctgtggccagagagaacaacccctggtgtgtacaccaaagt 1635
|||||
Db 1143 TGTCAACAGCTGGGGACAGGCTGTGGCCAGAGAAACAACTGGTGTGTACACCAAAAGT 1202
QY 1636 gacagagttctccctcgatttacagcaagatggagagcagagtgccgattca 1688
|||||
Db 1203 GACAGAAGTTCTCCCTGGATTTACAGCAAGATGGAGAGCGAGGTGCGATTCA 1255

RESULT 5
AXI190546/c 1314 bp DNA linear PAT 08-AUG-2001
LOCUS
DEFINITION
ACCESSION AXI190546
VERSION AXI190546.1 GI:15143857
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 1314)
Quinn, K.E., Spytak, K.A., Majumder, K., Vernet, C., Burgess, C.,
Fernandes, E., Taupier, R.J., Rastelli, L. and Herrmann, J.L.
TITLE Serine/threonine kinase and serine protease polypeptides and
nucleic acids encoding same
JOURNAL patent: WO 0136645-A 17 25-MAY-2001;
Curagen Corporation (US)
FEATURES
source Location/Qualifiers
1..1314
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 266 a 376 c 384 g 288 t
ORIGIN

Query Match 71.7%; Score 1253; DB 6; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 ggagagccaggtacagcctgcccaagttcacctggcgaggggccagaagcagctacc 495
|||||
Db 1312 GGAGAGCCAGGTACGAGCCTGCCAAGTTACCTGGCGGGAGGGCCAGAAGCAGTACC 1253
QY 496 gtcctcgggtgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 555
|||||
Db 1252 GCTATPCGGGTGCGTCTCCTCTCATTCGCTGTGGTGTTCGCTCATCATCTCTCTTCCA 1193
QY 556 gttctggcaggggccacacagggatcaggttacagagcagagagagagctgtcccaagca 615
|||||
Db 1192 GTTCTGGCAGGGCCACACAGGGATCAGGTACAAGGACGAGAGAGAGCTGTCCCAAGCA 1133
QY 616 cgcgtgtcgtgtgagcggggtggtgactgcaagctgaagagtgaagcagctgggctgcgt 675
|||||
Db 1132 CGCTGTTCGCTGTGACGGGGTGGTGACTGCAAGCTGAAAGTAGACGAGCTGGGCTGCGT 1073
QY 676 gagggttgactgggacaagtctctgttaaaatcactcctggttccctcccatcagtggt 735
|||||
Db 1072 GAGGTTTGACTGGGACAAGCTCTGCTTAAATCTACTCTGGGTCTCTCCATCAGTGGCT 1013
QY 736 tcccatctgacgacaactggaatgactccctactcagagagacacctgcagcagctggg 795
|||||
Db 1012 TCCCATCTGTACAGCAACTGGAATGACTCTTACTCAGAGAGAGACCTGCCAGCAGCTGG 953

QY 796 ttctgagagtgtcaccgcgacacaccgaggttgccccacagggattttgccaacagctcttc 855
Db 952 TTTTCGAGAGTGTCTACCGGACAAACCGAGGTTGCCACAGGGATTTTGCCAAACAGCTTCTC 893
QY 856 aatcttgagatacaaatccacatccaggaagaacctccacaggtctgaatgccttcccc 915
Db 892 AATCTTGTAGATACAACCTCCACCATCCAGGAAGCCCTCCACAGGCTGTGAATGCGCTTCCCA 833
QY 916 ggggtatatctccctccagtggttccctcactgcgagactgagggccatgacggcgagatcgt 975
Db 832 GGGGTATATCTCCTCCAGTGTTCCTCAGCTGCGGACCTGAGGGCCATGACCGGGCGGATCGT 773
QY 976 gggagggcgctggcctcgtagcaagtgggccttgccaagtgaagtctgacacttcggcac 1035
Db 772 GGGAGGGCGCTGGCCCTCGGATAGCAAGTGGCCTTGCAAGTGAGTCTGCACCTTCGCGCAC 713
QY 1036 caccacacatctgtggaggcacgctcattgacgcccagtggggtgctcactgcgcccactg 1095
Db 712 CACCCACATCTGTGGAGGACCGCTCATTCGCCCCAGTGGTGTCTCCTCCTGCCCCACTG 653
QY 1096 ctctctgtgacccgggagaaggtcctcgagggctggaaggtgtacgcgggaccagcaa 1155
Db 652 CTCTCTGTGTGACCCGGGAGAAAGTCTCTGGAGGGCTGGAAGGTGTAGCCGGGACACAGCAA 593
QY 1156 cctgcaccagttgctgagggcagcctccattgcccggagatcattcaacagcaattacac 1215
Db 592 CTGTCACCATCTGTGGAGGAGCCCTCCATTCGCCGAGATCATCATCAACAGCAATTTACAC 533
QY 1216 cgtatgaggaggagactatgacatcgccctcactgcggtgtccaaagccctgacctgtc 1275
Db 532 CGATGAGGAGGACGACTATGACATCGCCCTCATTCGCCGCTGTCCAAAGCCCTGACCCCTGTC 473
QY 1276 cgtcacatccacctgctgctcccccattgcatgacagacctttgacctcaatgagac 1335
Db 472 CGCTCACATCCACCTGCTTGGCTCCCATGATGATGACAGAGACCTTTAGCCTCAATGAGAC 413
QY 1336 ctgtggtatcacaggctttggcaagaccaggagagacagatgacaagacatccccctcct 1395
Db 412 CTGTGGATCACAGGCTTTGGCAAGACCAGGGAGACAGATGACAAGACATCCCCCTCCT 353
QY 1396 ccgggaggtgcaggtcaatctcactgactcaagaatacaatcaatgactactgtgtatga 1455
Db 352 CCGGGAGGTGCAGGTCATCTCATCGACTTCAAGAAATGCAATGACTTGTGTCTATGA 293
QY 1456 cagttaccttaccaccaaggtatgtgtctggggacctctgtggggcagagactccctg 1515
Db 292 CAGTTACCTTACCCCAAGATGATGTGCTGGGACCTTCGTGGGGCAGAGACTCCTG 233
QY 1516 ccaggagacagcgggggcccctctgtctgtgagcagaacaaccgctgttacctggcagg 1575
Db 232 CCAGGAGACAGCGGGGGCCCTTTGTCTGTGAGCAGAAACAACCCGCTGTGTACCCAAAGT 173
QY 1576 tctcaccagctgggcacagcgtgtggccagagaacaacctggtgtgtacaccaagt 1635
Db 172 GTGTACCACTGGGCGACAGGCTGTGGCCAGAGAAACAACCTGGTGTGTACCCAAAGT 113
QY 1636 gacagaagttcttccctgatttaccagcaagatggagagcagaggtgcgattca 1688
Db 112 GACAGAAGTTCTTCCTGGATTTACAGCAAGATGGAGAGCGAGGTGCGATTCA 60

RESULT 6
AB048797 3324 bp mRNA linear PRI 22-MAR-2001
LOCUS
DEFINITION Homo sapiens mspms mRNA for mosaic serine protease, complete cds.
ACCESSION AB048797
VERSION AB048797.1 GI:13429971
KEYWORDS
SOURCE Homo sapiens lung cDNA to mRNA, clone_lib:Marathon Ready cDNA (Clontech).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE	1 (sites)
AUTHORS	Kim,D.R., Sharmin,S., Inoue,M. and Kido,H.
TITLE	Cloning and expression of novel mosaic serine proteases with and without a transmembrane domain from human lung(1)
JOURNAL	Biochim. Biophys. Acta 1518 (1-2), 204-209 (2001)
MEDLINE	21167393
REFERENCE	2 (bases 1 to 3324)
AUTHORS	Kim,D.R., Inoue,M. and Kido,H.
TITLE	Direct Submission
JOURNAL	Submitted (12-SEP-2000) Hiroshi Kido, The University of Tokushima, Division of Enzyme Chemistry, Institute for Enzyme Research; Kuramoto-cho, Tokushima 770, Japan
	(E-mail: kido@ier.tokushima-u.ac.jp, Tel:81-88-633-7424, Fax:81-88-633-7425)

FEATURES	SOURCE
1. Age	1. Age
2. Gender	2. Gender
3. Marital Status	3. Marital Status
4. Education	4. Education
5. Income	5. Income
6. Occupation	6. Occupation
7. Religion	7. Religion
8. Political Affiliation	8. Political Affiliation
9. Health Status	9. Health Status
10. Travel History	10. Travel History
11. Employment Status	11. Employment Status
12. Family Size	12. Family Size
13. Home Ownership	13. Home Ownership
14. Vehicle Ownership	14. Vehicle Ownership
15. Insurance Status	15. Insurance Status
16. Subscription Services	16. Subscription Services
17. Charitable Donations	17. Charitable Donations
18. Volunteer Work	18. Volunteer Work
19. Real Estate Transactions	19. Real Estate Transactions
20. Financial Investments	20. Financial Investments
21. Medical History	21. Medical History
22. Legal Records	22. Legal Records
23. Criminal Record	23. Criminal Record
24. Academic Achievements	24. Academic Achievements
25. Professional Certifications	25. Professional Certifications
26. Travel Itineraries	26. Travel Itineraries
27. Hotel Reservations	27. Hotel Reservations
28. Airline Tickets	28. Airline Tickets
29. Rental Agreements	29. Rental Agreements
30. Utility Bills	30. Utility Bills
31. Bank Statements	31. Bank Statements
32. Credit Reports	32. Credit Reports
33. Tax Returns	33. Tax Returns
34. Insurance Policies	34. Insurance Policies
35. Subscription Cancellations	35. Subscription Cancellations
36. Charitable Receipts	36. Charitable Receipts
37. Volunteer Hours	37. Volunteer Hours
38. Real Estate Listings	38. Real Estate Listings
39. Investment Portfolios	39. Investment Portfolios
40. Medical Records	40. Medical Records
41. Legal Filings	41. Legal Filings
42. Criminal Records	42. Criminal Records
43. Academic Transcripts	43. Academic Transcripts
44. Professional Licenses	44. Professional Licenses
45. Travel Itineraries	45. Travel Itineraries
46. Hotel Reservations	46. Hotel Reservations
47. Airline Tickets	47. Airline Tickets
48. Rental Agreements	48. Rental Agreements
49. Utility Bills	49. Utility Bills
50. Bank Statements	50. Bank Statements
51. Credit Reports	51. Credit Reports
52. Tax Returns	52. Tax Returns
53. Insurance Policies	53. Insurance Policies
54. Subscription Cancellations	54. Subscription Cancellations
55. Charitable Receipts	55. Charitable Receipts
56. Volunteer Hours	56. Volunteer Hours
57. Real Estate Listings	57. Real Estate Listings
58. Investment Portfolios	58. Investment Portfolios
59. Medical Records	59. Medical Records
60. Legal Filings	60. Legal Filings
61. Criminal Records	61. Criminal Records
62. Academic Transcripts	62. Academic Transcripts
63. Professional Licenses	63. Professional Licenses
64. Travel Itineraries	64. Travel Itineraries
65. Hotel Reservations	65. Hotel Reservations
66. Airline Tickets	66. Airline Tickets
67. Rental Agreements	67. Rental Agreements
68. Utility Bills	68. Utility Bills
69. Bank Statements	69. Bank Statements
70. Credit Reports	70. Credit Reports
71. Tax Returns	71. Tax Returns
72. Insurance Policies	72. Insurance Policies
73. Subscription Cancellations	73. Subscription Cancellations
74. Charitable Receipts	74. Charitable Receipts
75. Volunteer Hours	75. Volunteer Hours
76. Real Estate Listings	76. Real Estate Listings
77. Investment Portfolios	77. Investment Portfolios
78. Medical Records	78. Medical Records
79. Legal Filings	79. Legal Filings
80. Criminal Records	80. Criminal Records
81. Academic Transcripts	81. Academic Transcripts
82. Professional Licenses	82. Professional Licenses
83. Travel Itineraries	83. Travel Itineraries
84. Hotel Reservations	84. Hotel Reservations
85. Airline Tickets	85. Airline Tickets
86. Rental Agreements	86. Rental Agreements
87. Utility Bills	87. Utility Bills
88. Bank Statements	88. Bank Statements
89. Credit Reports	89. Credit Reports
90. Tax Returns	90. Tax Returns
91. Insurance Policies	91. Insurance Policies
92. Subscription Cancellations	92. Subscription Cancellations
93. Charitable Receipts	93. Charitable Receipts
94. Volunteer Hours	94. Volunteer Hours
95. Real Estate Listings	95. Real Estate Listings
96. Investment Portfolios	96. Investment Portfolios
97. Medical Records	97. Medical Records
98. Legal Filings	98. Legal Filings
99. Criminal Records	99. Criminal Records
100. Academic Transcripts	100. Academic Transcripts

5' UTR
gene
CDS

```

/byte_mays
/note="non-transmembrane type"
/codon_start=1
/transl_except=(pos:1498..1500,aa:His)
/transl_except=(pos:787..789,aa:Arg)
/transl_except=(pos:904..906,aa:His)
/product="mosaic serine protease"

```

3'UTR	polyA_signal	BASE COUNT	ORIGIN
1702..3324	3273..3278	763 a 956 c 906 g 699 t	

```
Query Match      68.0%; Score 1188; DB 9; Length 3324;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

552	QY	tcagttctggaaggccacacagggatcaggtacaaggagcagagggagagctgtccca	611
553	DB		
554	DB	TCAGTTCTGGCAGGCCACACAGGATCAGGTACAAAGAGCAGAGGGAGAGCTGTCCA	613
555	DB		
612	QY	agcacgctgtcgctgtgacgggtggtggactgcaagctgaagatcacgactaggct	671
613	DB		
614	DB	AGCACGCTTTCGCTGTGACGGGGTGGTGGACTGCNAGCTGAAGAGTGACGAGCTGGGT	673
615	DB		
672	QY	ggatgaggtttgaactgggacaagtctctgcttaaaatctactctgggtctctcccatcagt	731
673	DB		
674	DB	GGGTGAGGTTTGACTGGGACAAGTCTCTGCTTAAATCTACTTGGTGCTCTCCCATCAGT	733
675	DB		
732	QY	ggcttcccatctgtagcagcaactgaaatgactcctactcagaagaacctgccagcagc	791
733	DB		
734	DB	GGCTTCCCATCTGTAGCACCACTGGAATGACTCTACTACAGAGAGACCTGCCAGCAGC	793
735	DB		
792	QY	tgggtttcagagtgctcaccggacaacacgaggttgccacagggattttgccaacagct	851
793	DB		
794	DB	TGGGTTTCAGAGTGCTCACCGGCACAACCGAGGTTGCCACAGGAGATTTTGCAACAGCT	853
795	DB		

Qy	852	tctcaatcttgagatacaactccacactccaggaagcctccacaggctcgaatgcctt	911
Db	853		912
Db	854	TCTCAATCTTGAGATACACTCCACCATCCAGGAAGCCTCCACAGGTCGTAATGCCCTT	913
Qy	912	ccagcgggtatctccctccagtgtccactcgcgactcgaggtcactgacggcgga	971
Db	913		972
Db	914	CCCAAGCGGTATATCTCCCTCCAGTGTCCCACTCGGACTGAGGCCATGACCGGGCGGA	973
Qy	972	tcgtggaggggcgctggcctcggatagcaagtggccttggcaagtgaagtgcacttcg	1031
Db	973		1032
Db	974	TCGTGGAGAGGGGCGTGGCTCGGATAGCAAGTGGCCTTGGCAAGTAGTGTGCACATCG	1033
Qy	1032	gcaccaccacatctggaagcacgctcatgacgcgccagtgcagccagtgggtgctcactccgcc	1091
Db	1033		1092
Db	1034	GCACCACCACATCTGTGGAGGCACGCTCATTTGACGCCAGTGGGTGCTCATCTCCGCC	1093
Qy	1092	actgcttcttcgtgacccgggagaaggtcctggagggtcggaaggtgtacgcgggacca	1151
Db	1093		1152
Db	1094	ACTGCTTCTTCTGAGCCGGAGAGAGGTCTCGAGGGGCTGGAAGGTGTACCGGGCACCA	1153
Qy	1152	gcaactgcacagttgcctgagcagcctcattgccgaatcatcatcaacgaatt	1211
Db	1153		1212
Db	1154	GCAACCTGTCAACAGTGTCCCTGAGCGAGCCTCCATTGCCAGATCATCATCAACAGCAAT	1213
Qy	1212	acccgatgaggagcagactatgacatcgccctcatcgctgtccaaagccctgaccc	1271
Db	1213		1272
Db	1214	ACCCGATGAGGAGGACGACTATGACATCGCCCTCATCGCGTGTCCAAGCCCTGACCC	1273
Qy	1272	tgtccgtcaatcaacacctgttcctcccacatgcattggacagacatttagacctgaatg	1331
Db	1273		1332
Db	1274	TGTCGGCTCACTACCAACCTGTCTGGCTCCCTCCCATGCATGGACAGACCTTTAGCCCTCAATG	1333
Qy	1332	agacctgtgattcaacaggtcttggcaagacaggagagacagatgacaagacatccccct	1391
Db	1333		1392
Db	1334	AGACCTGCTGATACAGGCTTTGGCAAGACCAGGAGACAGATGACAAAGACATCCCCCT	1393
Qy	1392	tcctcgggaggtgcaaggtcaatctcatcgacttcaagaaatgaatgaactacttgctct	1451
Db	1393		1452
Db	1394	TCCTCCGGGAGGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTTACTTGTGCT	1453
Qy	1452	atgacagttaccttaccctcccaaggatgatgtgctggggaccttcgtggggcgagagact	1511
Db	1453		1512
Db	1454	ATGACAGTTACTTACCCCAAGGATGATGTGCTTGGGGACCTTCGTGGGGGCGACAGACT	1513
Qy	1512	cctgcaggagagacagcggggggcctcttgtctgtgagcaaacacgcctggctacctgg	1571
Db	1513		1572
Db	1514	CCTGCCAGGGAGACAGCGGGGGCCTCTTGTGTGTGAGCAAGAACACCGCTGGTACCTGG	1573
Qy	1572	caagtgtaaccagctggggcacaggctgtggccagagaaaaaacacctgggtgtatacca	1631
Db	1573		1632
Db	1574	CAGGTGTACACAGCTGGGCGCAGGCTGTGGCCAGAGAAAACAAACCTGTGTGTACACCA	1633
Qy	1632	aagtgcacagaagttcttcctcctggatttacagcaagatggagcgaggtgcgattcaga	1691
Db	1633		1692
Db	1634	ANGTGACAGAAGTTCTTCCTGGATTACACCAAGATGGAGCGAGGTGCGATTACGAA	1693
Qy	1692	aatcctaaccagctggcctgctgctcgtgcacagcacccggctgctgtga	1739
Db	1693		1740
Db	1694	AATCCTTAACCACTGGCCCTGCTCTGTCAGACGACCGGGCTGCTGTGA	1741

RESULT 7

AK027798

LOCUS

DEFINITION

ACCESSION

ACCESSION
VERSION

KEYWORDS

SOURCE

ORGANIC

ORGANIS

AK027798 2192 bp mRNA linear PRI 15-MAY-2001
Homo sapiens cDNA FLJ14892 fis, clone PLACE1004270, weakly similar
to TRANSMEMBRANE PROTEASE, SERINE2 (EC 3.4.21.-).

AK027798

AK027798.1 GI:14042742

oligo capping; fis (ful

Homo sapiens placenta clone: PI.ACE1004270

CLONE: FLACET004270.
Homo sapiens

Eukaryota; Metazoa; Choro-

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (sites)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA
Published Only in DataBase (1999) In press
2 (bases 1 to 184926)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Feb 21, 2001 this sequence version replaced gi:11094157.
COMMENT
FEATURES
source
1..184926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="Rp11-728F11"
BASE COUNT 47261 a 47596 c 43757 g 46312 t
ORIGIN

Query Match 13.0%; Score 227; DB 9; Length 184926;
Best Local Similarity 100.0%; Pred. No. 1.6e-121;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 gaatgcattccaggaagaacattcagctgagcattccagccagcattctccagc 90
|||||
Db 163269 GAATGTCATCTCCAGCAGAACACCTTCAGCTGGAGCATCTCCAGCCAGCATCTCCAGC 163210
|||||

QY 91 tgggacacctccagcgccgcatctccagccagcattctccagccagcattctccagc 150
|||||
Db 163209 TGGGACACCTCCAGCGCGGCGCATCTCCAGCCAGCATCTCCAGCCAGCATCTCCAGC 163150
|||||

QY 151 tgggacacctccgccccgagcattccagccagcattctccagctggtcacctccagg 210
|||||
Db 163149 TGGGACACCTCCGCGGCGGCGCATCTCCAGCCAGCATCTCCAGCTGGTACACCTCCAGG 163090
|||||

QY 211 ccgggcatctccagcgccgagcattccagccagcattctccagcc 257
|||||
Db 163089 CCGGGCATCTCCAGCGGCGGCGCATCTCCAGCCAGCATCTCCAGCC 163043
|||||

RESULT 11
AP000683/C
LOCUS
DEFINITION Homo sapiens chromosome 11 clone CMB9-3G3 map 11q23, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION AP000683
VERSION AP000683.2 GI:8118871
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 111094)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 111,094 genomic DNA of 11q23
Published Only in DataBase (1999) In press
2 (bases 1 to 111094)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (08-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

COMMENT

Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997557.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-3G3
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 100488 bases at least Q40
Consensus quality: 105330 bases at least Q30
Consensus quality: 107871 bases at least Q20
Insert size: 109694; sum-of-contigs
Quality coverage: 4.61x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
15 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1	17437	contig of	17437	bp in length
17338	32179	contig of	14642	bp in length
32280	47896	contig of	15617	bp in length
47997	57580	contig of	9584	bp in length
57681	67202	contig of	9522	bp in length
67303	73887	contig of	6585	bp in length
73988	80352	contig of	6365	bp in length
80453	85964	contig of	5312	bp in length
86065	90832	contig of	4768	bp in length
90933	95398	contig of	4466	bp in length
95499	99811	contig of	4313	bp in length
99912	104226	contig of	4315	bp in length
104327	107335	contig of	3009	bp in length
107436	109226	contig of	1791	bp in length
109327	111094	contig of	1768	bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	17437	contig of	17437	bp in length
17438	17537	gap of	100	bp
17538	32179	contig of	14642	bp in length
32180	32279	gap of	100	bp
32280	47896	contig of	15617	bp in length
47897	47996	gap of	100	bp
47997	57580	contig of	9584	bp in length
57581	57680	gap of	100	bp
57681	67202	contig of	9522	bp in length
67203	67302	gap of	100	bp
67303	73887	contig of	6585	bp in length
73888	73987	gap of	100	bp
73988	80352	contig of	6365	bp in length
80453	80452	gap of	100	bp
80453	85964	contig of	5312	bp in length
85965	86064	gap of	100	bp
86065	90832	contig of	4768	bp in length
90833	90932	gap of	100	bp
90933	95398	contig of	4466	bp in length
95399	95498	gap of	100	bp

KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 450)
AUTHORS	Williams,L.F., Escobedo,J., Innis,M.A., Garcia,P.D., Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D., Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M., Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and Stache-Crain,B.
TITLE	Human genes and gene expression products
JOURNAL	Patent: WO 0166753-A 1591 13-SEP-2001;
FEATURES	Chiron Corporation (US) ; Hyseq Inc. (US) Location/Qualifiers source 1..450 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	97 a 166 c 113 g 74 t
ORIGIN	
Query Match	8.7%; Score 152; DB 6; Length 450;
Best Local Similarity	100.0%; Pred. No. 4e-77;
Matches 152; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy 1	ctcagagaccatgagagggacagccagcggaatgcatctccagcaagaacaccttcagc 60
Db 65	CTCAGAGACCATGGAGAGGGACAGCCACGGGAATGCATCTCCAGCAAGAACACCTTCAGC 124
Qy 61	tggagcatctccagccaggagcatctccagctgggagaccctccagggcgccatctccagc 120
Db 125	TGGAGCATCTCCAGCCAGGCATCTCCAGCTGGGACACCTCCAGGCGGGCATCTCCAGC 184
Qy 121	ccaggcatctccagccaggagcatctccagctg 152
Db 185	CCAGGCATCTCCAGCCAGGCATCTCCAGCTG 216
RESULT 13	
BC010843	
LOCUS	BC010843 2125 bp mRNA linear ROD 17-JUL-2001
DEFINITION	Mus musculus, Similar to mosaic serine protease, clone
IMAGE:3490022, mRNA, partial cds.	
ACCESSION	BC010843
VERSION	BC010843.1
KEYWORDS	GI:14789982
SOURCE	mouse, muscu

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 2125)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (12-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapps-remail.nih.gov Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 7 Row: b Column: 18

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES
source

Location/Qualifiers

1..2125
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="3490022"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCLCGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..531
/codon_start=1
/product="Similar to mosaic serine protease"
/protein_id="AAH10843.1"
/db_xref="GI:14789983"
/translation="PEAASISQIIINGNYTDDDDYDIALIRLSKPLTLSAHHPAACL
PMHGTFGLNETCWITGFKTKETDEKTSPEFLREYQVNLIDFKKNDYLVYDSYLT
NMAGDLRGGDRSCGGPLVCEQNNRWYLAGVTSWGTGGQKNKPGVYTKVTEVL
PWYRKMESEVFRKS"
536 a 502 c 541 g 546 t

CDS

Query Match 2.8%; Score 49; DB 10; Length 2125;

Best Local Similarity 100.0%; Pred. No. 1.5e-16;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1559 cgcgtgtaactgcaggtgtccacagctggggcacaggtgtggccaga 1607

|||||
Db 391 CGCTGTACTGGCAGGTGTACACAGCTGGGGCACAGGCTGTGGCCAGA 439

RESULT 14

AX149587/c

LOCUS AX149587 29 bp DNA linear PAT 08-JUN-2001

DEFINITION Sequence 11 from Patent WO0136604.

ACCESSION AX149587

VERSION AX149587.1 GI:14348021

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 29)

AUTHORS Madison, E.L. and Ong, E.O.

TITLE Nucleic acids encoding endotheliases, endotheliases and uses

thereof

JOURNAL Patent: WO 0136604-A 11 25-MAY-2001;

CORVAS INTERNATIONAL, INC. (US)

FEATURES

source

1..29

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Oligonucleotide Primer"

7 a 4 c 15 g 3 t

BASE COUNT

ORIGIN

Query Match 1.7%; Score 29; DB 6; Length 29;

Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1273 gtccgctcacatccaccctgtgtgctcc 1301

|||||
Db 29 GTCCGCTCACATCCACCCCTGCTTGCTCC 1

RESULT 15

AX149595/c

LOCUS AX149595 29 bp DNA linear PAT 08-JUN-2001

DEFINITION Sequence 19 from Patent WO0136604.

ACCESSION AX149595

VERSION AX149595.1 GI:14348028

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 29)

AUTHORS Madison, E.L. and Ong, E.O.

TITLE Nucleic acids encoding endotheliases, endotheliases and uses

thereof

JOURNAL Patent: WO 0136604-A 19 25-MAY-2001;

CORVAS INTERNATIONAL, INC. (US)

FEATURES

source

1..29

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Oligonucleotide primer"

7 a 4 c 15 g 3 t

BASE COUNT

ORIGIN

Query Match 1.7%; Score 29; DB 6; Length 29;

Best Local Similarity 100.0%; Pred. No. 0.00014;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1273 gtccgctcacatccaccctgtgtgctcc 1301

|||||

Db 29 GTCCGCTCACATCCACCCCTGCTTGCTCC 1

Search completed: August 19, 2002, 23:51:52

Job time: 10779 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 19:49:08 ; Search time 291.64 Seconds
(without alignments)
10290.654 Million cell updates/sec

Title: US-09-879-792-11
Perfect score: 1748
Sequence: 1 ctcagagaccatgagagagg.....ggctgctgtgactcgagaaa 1748

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1678	96.0	1689	22 AAD05796	Human transmembran
2	1663	95.1	2067	22 AAD05797	Human transmembran
3	1253	71.7	1314	22 AAF83971	Human SER6 nucleot
c 4	1253	71.7	1314	22 AAF83973	Nucleotide sequenc
5	1049	60.0	2192	22 AAH14850	Human cDNA sequenc
c 6	1019	58.3	1078	22 AAF83972	Nucleotide sequenc
7	782	44.7	1222	21 AAD02322	Human serine prote
8	257	14.7	852	22 AAH07577	Human cDNA clone (
9	250	14.3	1212	23 AAS91044	DNA encoding novel

10	156	8.9	926	22	AAI02725	Human reproductive
11	152	8.7	450	22	AAS38533	Novel human diagno
12	146	8.4	705	23	AAS79452	DNA encoding novel
13	145	8.3	926	22	AAI02724	Human reproductive
c 14	29	1.7	29	22	AAD05802	Human endotheliase
c 15	29	1.7	29	22	AAD05809	Human endotheliase
16	27	1.5	27	22	AAD05804	Human endotheliase
17	27	1.5	27	22	AAD05805	Human endotheliase
18	26	1.5	26	22	AAD05803	Human endotheliase
19	25	1.4	25	22	AAD05810	Human endotheliase
c 20	25	1.4	36	22	AAD05806	Human endotheliase
21	24	1.4	39	22	AAD05815	Human endotheliase
22	24	1.4	509	23	AAS87565	DNA encoding novel
c 23	24	1.4	969	22	AAD14979	Human Stratum corn
24	24	1.4	986	16	AAQ81203	Human amyloid prec
c 25	24	1.4	1089	17	AAT39783	SER7 gene specific
26	23	1.3	23	22	AAF83975	SER7 gene specific
27	22	1.3	22	22	AAD05808	Human endotheliase
28	22	1.3	1482	23	ABL27021	Drosophila melanog
29	22	1.3	3630	23	ABL27020	Drosophila melanog
c 30	21	1.2	21	22	AAF83974	SER7 gene specific
31	21	1.2	21	22	AAF83976	SER7 gene specific
32	21	1.2	4707	20	AAK55560	Viral-encoded sema
33	21	1.2	4707	21	AAC63721	Human VESPR cDNA.
34	21	1.2	4707	22	AAS00320	Viral-encoded sema
35	21	1.2	4707	22	AAS00459	Human VESPR cDNA.
36	21	1.2	4707	22	AAF76945	Nucleotide sequenc
37	21	1.2	4707	22	AAF25249	Human secreted exp
c 38	20	1.1	251	21	AAA45082	CNA encoding mous
39	20	1.1	717	21	AAA61659	CNA encoding mous
40	20	1.1	1685	21	AAA61660	CNA encoding mous
41	20	1.1	2068	21	AAA61661	CNA encoding mous
42	20	1.1	2070	21	AAA61662	CNA encoding mous
43	19	1.1	631	21	AAC64228	Human pollinosis-a
44	19	1.1	1030	21	AAC64229	Human pollinosis-a
45	19	1.1	1034	21	AAA61559	CNA encoding mous

ALIGNMENTS

RESULT 1

AAD05796

ID AAD05796 standard; DNA; 1689 BP.

AC AAD05796;

XX 31-JUL-2001 (first entry)

DT Human transmembrane serine protease (Endotheliase 2-S) DNA.

DE Human; endotheliase 2-S; protease domain; cytosolic; vulnary; wound;

KW neotropic; periodontitis; dermatological disorder; psoriasis; neovascular disease;

KW angioedema; cardiovascular disorder; psoriasis; neovascular disease;

KW chronic inflammatory disease; ocular disorder; circulatory disorder;

KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;

KW liver cirrhosis; osteoradionecrosis; systemic sclerosis; oesophageal;

KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;

KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;

KW transmembrane serine protease; ds.

XX Homo sapiens.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

FT CDS

FT 1..1689

FT /*tag= a

FT /product= "Human endotheliase 2-S protein"

FT WO200136604-A2.

XX 25-MAY-2001.

PD 17-NOV-2000; 2000WO-US31803.

XX DNA encoding novel

PF

18-NOV-1999; 990US-0166391.
22-SEP-2000; 2000US-0234840.
(CORV-) CORVAS INT INC.
Madison EL, Ong EO;
WPI; 2001-336001/35.
P-PSDB; AAE01943.
New nucleic acid encoding a protein comprising endotheliase activity
useful in the prevention and treatment of e.g. vascular malformations,
cardiovascular disorders, and chronic inflammatory disease -
Example 2; Page 135-137; 152pp; English.
The present DNA sequence encodes human short form transmembrane serine
protease (Endotheliase 2-S) protein.
The invention relates to an endotheliase protein, endotheliase protease
domain and their corresponding nucleic acid molecules. An endotheliase
protein or protease domain of it is useful for the treatment and
diagnosis of disorders associated with aberrant angiogenesis or undesired
neovascularisation. The undesired angiogenesis is associated with
disorders selected from solid neoplasm, vascular malformations and
cardiovascular disorders such as angiofibroma, angiolipoma,
atherosclerosis, restenosis/reperfusion injury, arteriovenous
malformations, haemangiomatosis and vascular adhesions, dyschondroplasia
with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
syndrome, chronic inflammatory diseases such as diabetes mellitus,
haemophilic joints, inflammatory bowel disease, nonhealing fractures,
periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
granulations-burns, hypertrophic scars, liver cirrhosis,
osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
systemic sclerosis and aberrant wound repairs, circulatory disorders
Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
dyomyotility, sclerodactyly and teangiectasis, dermatological disorders
such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, port-wine
stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
caused by ocular neovascular disease, corneal graft neovascularisation,
macular degeneration, retinopathy of prematurity, retrolental
fibroplasia and corneal neovascularisation. The nucleic acids of the
invention are also used in gene therapy. The invention also provides
method for screening compounds that modulate angiogenesis.

```
Query Match          96.0%; Score 1678; DB 22; Length 1689;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	11	atgagaggagacgccacgggaatgcatctccagcaagaacaccttcagctggagcatct	70
Db	1	atggagaggagacgccacgggaatgcatctccagcaagaacaccttcagctggagcatct	60
Qy	71	ccagccaggcatctcagctgggacacctccaggcgggcatctccagccaggcatct	130
Db	61	ccagccaggcatctcagctgggacacctccaggcgggcatctccagccaggcatct	120
Qy	131	ccagccaggcatctcagctgggacacctccgggcgggcatctccagccaggcatct	190
Db	121	ccagccaggcatctcagctgggacacctccgggcgggcatctccagccaggcatct	180
Qy	191	ccagctggtaacctccaggccgggcatctccaggcgggcatctccagccaggcatct	250
Db	181	ccagctggtaacctccaggccgggcatctccaggcgggcatctccagccaggcatct	240
Qy	251	ccagcccgggcatctcggctcgtggcatcattccaggctcctcatccggcaggcatca	310

QY 1391 ttctccggagggtcaggtcgaatctcatcgaactcaagaatgcaatgacttggtc 1450
|||||
Db 1381 ttctccggagggtcaggtcgaatctcatcgaactcaagaatgcaatgacttggtc 1440
|||||
QY 1451 tatgacagttacattcccccaagatgatgtgctgctggggacattcgtggggcagagac 1510
|||||
Db 1441 tatgacagttacattcccccaagatgatgtgctgctggggacattcgtggggcagagac 1500
|||||
QY 1511 tctgcagggagacagcggggctcttctgtgagcagacaacacgctgtacctg 1570
|||||
Db 1501 tctgcagggagacagcggggctcttctgtgagcagacaacacgctgtacctg 1560
|||||
QY 1571 gcaggtgtcaccagctggggcagagctgtggcagagaaacaaacctgtgtgtacac 1630
|||||
Db 1561 gcaggtgtcaccagctggggcagagctgtggcagagaaacaaacctgtgtgtacac 1620
|||||
QY 1631 aaagtacagaattttctccctgatttacagcaaatgagagcgaggtgcattca 1688
|||||
Db 1621 aaagtacagaattttctccctgatttacagcaaatgagagcgaggtgcattca 1678
|||||

RESULT 2

AAD05797

ID AAD05797 standard; DNA; 2067 BP.

XX AAD05797;

XX AAD05797;

DT 31-JUL-2001 (first entry)

XX Human transmembrane serine protease (Endotheliase 2-L) DNA.

XX Human; endotheliase 2-L; protease domain; cytosstatic; vulnery; wound;

KW nontropic; periodontitis; dermatological disorder; gene therapy; scar;

KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;

KW chronic inflammatory disease; ocular disorder; circulatory disorder;

KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;

KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;

KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;

KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;

KW transmembrane serine protease; ds.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

CC The invention relates to an endotheliase protein, endotheliase protease
CC domain and their corresponding nucleic acid molecules. An endotheliase
CC protein or protease domain of it is useful for the treatment and
CC diagnosis of disorders associated with aberrant angiogenesis or undesired
CC neovascularisation. The undesired angiogenesis is associated with
CC disorders selected from solid neoplasm, vascular malformations and
CC cardiovascular disorders such as angiofibroma, angiolipoma,
CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
CC malformations, haemangiomas and vascular adhesions, dyschondroplasia
CC with vascular hamartomas (Pafucci's syndrome), hereditary haemorrhagic
CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
CC granuloma-in-burns, hypertrophic scars, liver cirrhosis,
CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
CC systemic sclerosis, and aberrant wound repairs, circulatory disorders
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
CC dyonotility, sclerodactyly and teagiectasis, dermatological disorders
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
CC caused by ocular neovascular disease, corneal graft neovascularisation,
CC macular degeneration, retinopathy of prematurity, retrolental
CC fibroplasia and corneal neovascularisation. The nucleic acids of the
CC invention are also used in gene therapy. The invention also provides
CC method for screening compounds that modulate angiogenesis.
CC
XX
SQ Sequence 2067 BP; 463 A; 604 C; 540 G; 460 T; 0 other;

Query Match 95.1%; Score 1663; DB 22; Length 2067;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 atggagaggagcagccagcgggaatgcatctccagcaagaacaccttcagctggagcatct 70
|||||
Db 1 atggagaggagcagccagcgggaatgcatctccagcaagaacaccttcagctggagcatct 60
|||||
QY 71 ccagcccgagcattccagctgggacacctccagccgggcatctccagccagggcatct 130
|||||
Db 61 ccagcccgagcattccagctgggacacctccagccgggcatctccagccagggcatct 120
|||||
QY 131 ccagcccgagcattccagctgggacacctccagccgggcatctccagccagggcatct 190
|||||
Db 121 ccagcccgagcattccagctgggacacctccagccgggcatctccagccagggcatct 180
|||||
QY 191 ccagctggtacacctccagccgggcatctccagccgggcatctccagccagggcatct 250
|||||
Db 181 ccagctggtacacctccagccgggcatctccagccgggcatctccagccagggcatct 240
|||||
QY 251 ccagcccgagcattccagctgggcatctccagcttcagcttcagcttcagcttcagct 310
|||||
Db 241 ccagcccgagcattccagctgggcatctccagcttcagcttcagcttcagcttcagct 300
|||||
QY 311 tccgccagctcagctcggtgacacccctcccaacagagtgatctgttagaacaaca 370
|||||
Db 301 tccgccagctcagctcggtgacacccctcccaacagagtgatctgttagaacaaca 360
|||||
QY 371 ccagtgggggctgtaccatccgatctctcctcagcttcagcttcagcttcagcttcagct 430
|||||
Db 361 ccagtgggggctgtaccatccgatctctcctcagcttcagcttcagcttcagcttcagct 420
|||||
QY 431 accagggagagccaggttacagcctgccaaagttcaactggcggggagggcagagcag 490
|||||
Db 421 accagggagagccaggttacagcctgccaaagttcaactggcggggagggcagagcag 480
|||||
QY 491 ctaccgctcagctcggtgctcctcctcattccctcctcctcctcctcctcctcctcctc 550
|||||
Db 481 ctaccgctcagctcggtgctcctcctcattccctcctcctcctcctcctcctcctcctc 540
|||||
QY 551 ttccagttctgaggggcccacacagggatcaggtacaagagagagagagagagagagagctgtccc 610
|||||

Db 541 ttccagttctcggcaggccacacaggtcaggtacaaggagcagagagagctgtccc 600
Qy aagcacgctgttcgctgtgacggggtggtggaactcaagctgaagctgacgagctgggc 670
Db 601 aagcacgctgttcgctgtgacggggtggtggaactcaagctgaagctgacgagctgggc 660
Qy tgcgtgaggtttgactgggagcaagtctctcttaaaatctactctgggtctcccatcag 730
Db 661 tgcgtgaggtttgactgggagcaagtctctcttaaaatctactctgggtctcccatcag 720
Qy tggcttcccatctgtacagcaactggaatgactctctactcagagaagacctgccagcag 790
Db 721 tggcttcccatctgtacagcaactggaatgactctctactcagagaagacctgccagcag 780
Qy ctgggtttcagagagctccaccgacacagaggttgcacacagagatttgcacacagc 850
Db 781 ctgggtttcagagagctccaccgacacagaggttgcacacagagatttgcacacagc 840
Qy ttctcaatcttgagatacaactccaccatccaggaaagcctccacaggtctgaatgccct 910
Db 841 ttctcaatcttgagatacaactccaccatccaggaaagcctccacaggtctgaatgccct 900
Qy tccacagcgttatatctccctccaggttcccaactcggactgagggccatgacccggcgg 970
Db 901 tccacagcgttatatctccctccaggttcccaactcggactgagggccatgacccggcgg 960
Qy atcgtggagggcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1030
Db 961 atcgtggagggcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1020
Qy ggcacacccacatctgtgagcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1090
Db 1021 ggcacacccacatctgtgagcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1080
Qy cactgcttctcgtgacccgggagaggtctcgtgagggctggaaggtgtacgcgccacc 1150
Db 1081 cactgcttctcgtgacccgggagaggtctcgtgagggctggaaggtgtacgcgccacc 1140
Qy agcaacctgcacagttgctgagcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1210
Db 1141 agcaacctgcacagttgctgagcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1200
Qy tacacgatgagagagcagctatgacatcgccctcatcgcgctgtccaaagccctgacc 1270
Db 1201 tacacgatgagagagcagctatgacatcgccctcatcgcgctgtccaaagccctgacc 1260
Qy ctgtcgcgtcacatccacctgcttgcctcccatgcatgagcagaccttagcttcaat 1330
Db 1261 ctgtcgcgtcacatccacctgcttgcctcccatgcatgagcagaccttagcttcaat 1320
Qy gagacctgctggatcacagcgtttgccaagaccagagagacagatgacaagacatcccc 1390
Db 1321 gagacctgctggatcacagcgtttgccaagaccagagagacagatgacaagacatcccc 1380
Qy ttctcgcggaggtgcaggtcaatctcctcagcttcaagaaatgcgaatgactacttggtc 1450
Db 1381 ttctcgcggaggtgcaggtcaatctcctcagcttcaagaaatgcgaatgactacttggtc 1440
Qy tatgacagttacattaccacaaagatgagtgtcgtcgtgggacacctcgtggggcgacagac 1510
Db 1441 tatgacagttacattaccacaaagatgagtgtcgtcgtgggacacctcgtggggcgacagac 1500
Qy tctcgcagggagacagcggggcctcttctgtcgtgagcagacaacacgctggtaccttg 1570
Db 1501 tctcgcagggagacagcggggcctcttctgtcgtgagcagacaacacgctggtacctg 1560
Qy gcagggtgtaccacagctggggcacagcgtgtggccagagaaaacacacgtggtgtacacc 1630
Db 1561 gcagggtgtaccacagctggggcacagcgtgtggccagagaaaacacacgtggtgtacacc 1620
Qy aaagtacagaagttctcctcgtgatttacagcaagatggaga 1673

Db 1621 aaagtacagaagttctcctcgtgatttaccagcaagatggaga 1663
RESULT 3
AAF83971
ID AAF83971 standard; DNA; 1314 BP.
XX
AC AAF83971;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human SER6 nucleotide sequence.
XX
KW Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
KW immunosuppressive; antiasthmatic; antifibrinolytic; SER6; ds.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /*tag= a
FT /product= "SER6"
XX
PN WO200136645-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31744.
XX
PR 17-NOV-1999; 99US-0165986.
PR 09-FEB-2000; 2000US-0181347.
PR 03-APR-2000; 2000US-0194195.
PR 05-APR-2000; 2000US-0194839.
PR 07-APR-2000; 2000US-0195637.
PR 13-APR-2000; 2000US-0197080.
PR 15-SEP-2000; 2000US-0232677.
PR 16-NOV-2000; 2000US-0165986.
XX
(CURA-) CURAGEN CORP.
XX
XX Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Taupler R, Rastelli L, Herrmann JL;
XX
DR WPI; 2001-336006/35.
DR P-PSDB; AAB85042.
XX
PT New isolated SERX polynucleotides and polypeptides related to the
serine/threonine kinase family of proteins, useful for determining and
treating diseases related to altered levels of the polypeptide -
XX
PS Claim 9; Page 33; 127pp; English.
XX
CC The invention relates to serine/threonine kinase-like proteins, termed
SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
via a recombinant expression vector in a host cell in gene therapy
applications), to detect SERX mRNA (e.g. in a biological sample) or a
genetic lesion in a SERX gene and modulate SERX activity. The SERX
polypeptides, polynucleotides, antibodies can also be used to treat or
prevent a pathology associated with SERX. Disorders associated with
aberrant SERX expression or activity include breast or testicular cancer,
cardiovascular defect, autoimmune disorders and asthma. Disorders
associated with aberrant SER5-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
represents the nucleotide sequence of human SER6.
XX
SQ Sequence 1314 BP; 288 A; 384 C; 376 G; 266 T; 0 other;

Query Match 71.7%; Score 1253; DB 22; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2192 BP; 531 A; 655 C; 536 G; 470 T; 0 other;

Query Match 60.0%; Score 1049; DB 22; Length 2192;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	229	ggcatctccagccagcagcattccagcccgccggtccggtctccggtctggcatcactttccag	288
DB	324	ggcatctccagccagcagcattccagcccgccggtccggtctggcatcactttccag	383
QY	289	gtctcatccggcaggttcattccatcccgccaggtcagctcgtgacacaccccccacaccag	348
DB	384	gtctcatccggcaggttcattccatcccgccaggtcagctcgtgacacaccccccacaccag	443
QY	349	agtgtacctgttagagcaaacacagtgagggtgtgacctatccgcatcctctctgccag	408
DB	444	agtgtacctgttagagcaaacacagtgagggtgtgacctatccgcatcctctctgccag	503
QY	409	gtcagcaccagcaaccagggccaccagggagagccaggttacagcctgcccaggttcac	468
DB	504	gtcagcaccagcaaccagggccaccagggagagccaggttacagcctgcccaggttcac	563
QY	469	ctggcgggagggcagagcagtcacgtcctatccgtggtgctcctctctctctccct	528
DB	564	ctggcgggagggcagagcagtcacgtcctatccgtggtgctcctctctctctccct	623
QY	529	ggtggttctcgtctatcctctctcagttctgagcggccacacagggatcaggtacaa	588
DB	624	ggtggttctcgtctatcctctctcagttctgagcggccacacagggatcaggtacaa	683
QY	589	ggagcagagggagagctgtcccaagcacgctgtctgctgagcggggtggtgactgcaa	648
DB	684	ggagcagagggagagctgtcccaagcacgctgtctgctgagcggggtggtgactgcaa	743
QY	649	gctgaagagtgaacagctgggtgctgaggtttgactgggacaaagtctctgttctaaat	708
DB	744	gctgaagagtgaacagctgggtgctgaggtttgactgggacaaagtctctgttctaaat	803
QY	709	ctactctgggtctcccatcagtggtctccatctgtagcagcaactgaaatcactcta	768
DB	804	ctactctgggtctcccatcagtggtctccatctgtagcagcaactgaaatcactcta	863
QY	769	ctcagaagaagacctgcagcagctgggtttcagagagtgtccaccggacacacaggttgc	828
DB	864	ctcagaagaagacctgcagcagctgggtttcagagagtgtccaccggacacacaggttgc	923
QY	829	ccacagggatttggcacaagcttctcaatcttgagatacaactccaccatccaggaag	888
DB	924	ccacagggatttggcacaagcttctcaatcttgagatacaactccaccatccaggaag	983
QY	889	ctccacaggtctgaatgaccttccagcgggtatctcctccaggtgttccactgcgg	948
DB	984	ctccacaggtctgaatgaccttccagcgggtatctcctccaggtgttccactgcgg	1043
QY	949	actgagggccatgacccggcgagctgtggagggcgctggcctcgatagcaagtggcc	1008
DB	1044	actgagggccatgacccggcgagctgtggagggcgctggcctcgatagcaagtggcc	1103
QY	1009	ttggcaagtgtgctcacttgcggcaccaccacacatctgtggaggcacctcattgacgc	1068
DB	1104	ttggcaagtgtgctcacttgcggcaccaccacacatctgtggaggcacctcattgacgc	1163
QY	1069	ccagtggtgtcactgcgcgcacactgttctctgtagccgggagagaggtctctgaggg	1128
DB	1164	ccagtggtgtcactgcgcgcacactgttctctgtagccgggagagaggtctctgaggg	1223

QY	1129	ctggaaggtgtacgcggccaccagcaacctgcaccagttgcctgagcagcctccattgc	1188
DB	1224	ctggaaggtgtacgcggccaccagcaacctgcaccagttgcctgagcagcctccattgc	1283
QY	1189	cgagatcatcatcaacagcaattacacccgatgagggagcagactatgacatcgccctcat	1248
DB	1284	cgagatcatcatcaacagcaattacacccgatgagggagcagactatgacatcgccctcat	1343
QY	1249	gcggtgttccaaagccctgacctgtccg	1277
DB	1344	gcggtgttccaaagccctgacctgtccg	1372

RESULT 6
AAF83972/c
ID AAF83972 standard; DNA; 1078 BP.
XX
AC AAF83972;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of SER7, a partial reverse complement of SER6.
XX
KW Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
KW cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
KW immunosuppressive; antiasthmatic; antifibrinolytic; SER6; SER7; ss.
XX
OS Homo sapiens.
XX
PN WO200136645-A2
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31744.
XX
PR 17-NOV-1999; 99US-0165986.
PR 09-FEB-2000; 2000US-0181347.
PR 03-APR-2000; 2000US-0194195.
PR 05-APR-2000; 2000US-0194839.
PR 07-APR-2000; 2000US-0195637.
PR 13-APR-2000; 2000US-0197080.
PR 15-SEP-2000; 2000US-0232677.
PR 16-NOV-2000; 2000US-0165986.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Tauplier R, Rastelli L, Herrmann JL;
XX
WPI; 2001-336006/35.
XX
PT New isolated SERX polynucleotides and polypeptides related to the
PT serine/threonine kinase family of proteins, useful for determining and
PT treating diseases related to altered levels of the polypeptide -
XX
PS Claim 9; Page 37; 127pp; English.
XX
CC The invention relates to serine/threonine kinase-like proteins, termed
CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
CC via a recombinant expression vector in a host cell in gene therapy
CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
CC polypeptides, polynucleotides, antibodies can also be used to treat or
CC prevent a pathology associated with SERX. Disorders associated with
CC aberrant SERX expression or activity include breast or testicular cancer,
CC cardiovascular defect, autoimmune disorders and asthma. Disorders
CC associated with aberrant SER5-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
CC represents the nucleotide sequence of human SER7, a partial reverse
XX complement of SER6.
SQ Sequence 1078 BP; 221 A; 296 C; 319 G; 242 T; 0 other;

Query Match		58.3%;	Score 1019;	DB 22;	Length 1078;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1019;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	670	ctgcgtgaggttgaactggaacagctctctgcttaaaatctactctctggtctccccaatca	729		
Db	1078	CTGCGTGAGGTTTGACTGGGACAGTCTCGCTTAAATCTACTCTGGTCTCCCATCA	1019		
Qy	730	gtggtctcccatctgtagcagcaactggaatgactctactctcagagaagacctgccagca	789		
Db	1018	GTGGCTTCCCATCTGTAGACAGCAACTGGAATGACTCTCTCAGAGAAGACCTGCCAGCA	959		
Qy	790	gctggggtttcgagagtgctcaccggacaaccgaggttgccacagggatctttgccaacag	849		
Db	958	GCTGGGTTTCGAGAGTGCTCACCGGACAACCGAGGTGGCCACAGGGGATTTGCCAACAG	899		
Qy	850	cttctcaatttgagatacaactccaccatccagggaagcctccacaggtctgaatgcc	909		
Db	898	CTTCTCAATCTTGAGATACAACTCCACATCCAGGAAAGCCTCCACAGGTCTGAATGCC	839		
Qy	910	ttcccagcggtatctctccctccagtggttccactgcggactgagggccatgaccggcg	969		
Db	838	TTCCACGGGTATATCTCCCTCCAGTGTCCACTGCGGACTGAGGCCCATGACCGGGCG	779		
Qy	970	gatgtggggggcgctggcctcggatagcaagtggccttggcaagtgaagtctgcactt	1029		
Db	778	GATCGTGGAGGGCGCTGGCCTCGGATACGAAGTGGCCCTGGCAAGTGAAGTCTGCACCT	719		
Qy	1030	cggcacaccacatctgtgagggcacgctcattgagcccgactgggtgctcactgccgc	1089		
Db	718	CGGCACACCCACATCTGTGGAGGCGAGCTCATTTGAGCCCACTGGGTGCTCACTGCCGC	659		
Qy	1090	ccactgcttctcgtgaaccggaggaagctcctggagggtcgaaggtgtacgcggcac	1149		
Db	658	CCACTGCTTCTTGTGNACCCGGAGAGGTCTCTGGAGGGCTGGAAGGTGTACCGGGGCAC	599		
Qy	1150	cagcaacctgcacagttgcctgaggcagcctccattgccgagatcatcatcaacagcaa	1209		
Db	598	CAGCAACCTGCACCAAGTTGGCTGAGGCGAGCTCCATTTGCCGAGATCATCAACAGCAA	539		
Qy	1210	ttacacgatgaggagagcaactatgacatcgccctcatcgctgcgaagccctgac	1269		
Db	538	TTACACCGATGAGGAGGACGACTATGACATCGGCCCTCATCGGCTGTCCAAAGCCCCGTGAC	479		
Qy	1270	cctgtccgtccacatccacccctgtctgcctcccccatagtcatggagacacctttagcctcaa	1329		
Db	478	CCTGTCCGCTCACATCCACCCCTGCTTGCTTCCCTCCCATGTCATGGACAGACCTTTAGCCTCAA	419		
Qy	1330	tgagacctgctgatacacagcgttctggcaagaccagggagacagatgacaagacatcccc	1389		
Db	418	TGAGACCTGTGGATCACAGGCTTTGGCAAGACGACGAGGAGACATCACAAAGACATCCCC	359		
Qy	1390	cttccctcggagaggtgcagctcaatctcatcgacttcaagaatgcaatgactacttgt	1449		
Db	358	CTTCCCTCCGGAGGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTACTTGGT	299		
Qy	1450	ctatgacagttaccttaccaccaaggatgatgtgtcgtgggaccttcgtgggggcagaga	1509		
Db	298	CTATGACAGTTACCTTACCCCAAGGATGATGTGTGTGGGGACCTTCGTGGGGGCAGAGA	239		
Qy	1510	ctcctgcaggagagacagcgggggocctctgtctgtgagcagaaacaacgcctggtacct	1569		
Db	238	CTCTCTGCCAGGAGACAGCGGGGGGCTCTTGTCTGTGTGAGCAGAAACAACCGCTGGTACCT	179		
Qy	1570	ggcaggtgtccacagctggggcacaggtgtgtgcccagagaacaacacctgtgtgtacac	1629		
Db	178	GGCAGGTGTCAACAGCTGGGGGCACAGCTGTGGCCAGAGAAACAACCTGGTGTGTACAC	119		
Qy	1630	caaaagtacagaagttcttccctggatttacacaaagatgagagcgaggtgcgattca	1688		
Db	118	CAAAAGTGACAGAAGTCTTCCCTGGATTATACCAAGATGGAGAGCGAGGTGGCATCA	60		

RESULT	7	
AAD02322		
ID	AAD02322 standard; cDNA; 1222 BP.	
XX		
AC	AAD02322;	
XX		
DT	28-MAR-2001 (first entry)	
XX		
DE	Human serine protease #3, encoding HATEE38 cDNA clone.	
XX		
KW	Human; serine protease; osteopathic; immunosuppressive; antiallergic;	
KW	antiinflammatory; cytostatic; cardiant; neuroprotective; nootropic;	
KW	neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;	
KW	antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;	
KW	treatment; bone formation disorder; osteoporosis; arthritis; cancer;	
KW	connective tissue disorder; autoimmune disorder; wound healing; asthma;	
KW	systemic lupus erythematosus; male reproductive system disorder;	
KW	testicular cancer; digestion and food absorption disorder; arrhythmia;	
KW	Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;	
KW	behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;	
KW	cardiovascular disorder; ocular disorder; drug screening; ss.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
CDS	76..1194	
FT	/tag= a	
FT	/product= "Human serine protease #3 from clone HATEE38"	
FT	76..963	
FT	/tag= b	
FT	/product= "Human serine protease #3 fragment from clone HATEE38"	
FT	/note= "Does not include stop codon"	
FT	/partial	
XX		
PN	WO200068247-A2.	
XX		
PD	16-NOV-2000.	
XX		
PF	05-MAY-2000; 2000WO-US12207.	
XX		
PR	07-MAY-1999; 99US-0133239.	
PR	20-MAY-1999; 99US-0135163.	
PR	03-AUG-1999; 99US-0147005.	
PR	09-SEP-1999; 99US-0152935.	
PR	01-NOV-1999; 99US-0162979.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Ruben SM, Shi Y, Young PE, Ni J;	
XX		
DR	WPI; 2000-679799/66.	
DR	P-PSDB; AAV72092, AAV72108.	
XX		
PT	New nucleic acid molecules encoding human serine protease polypeptides,	
PT	useful for diagnosis, prevention and/or treatment of disorders e.g.	
PT	osteoporosis, lupus erythematosus and Alzheimer's .	
XX		
PS	Claim 1; Page 271-272; 289pp; English.	
XX		
CC	The present cDNA sequence encodes human serine protease #3 from clone	
CC	HATEE38 (ATCC Deposit No: PTA27).	
CC	The invention relates to human serine proteases and their cDNA clones.	
CC	It is used in methods for the diagnosis, prevention and treatment of	
CC	various disorders related to serine protease such as bone formation	
CC	disorders (osteoporosis), connective tissue disorders (arthritis),	
CC	autoimmune disorders (systemic lupus erythematosus), wound healing, male	
CC	reproductive system disorders (testicular cancer), digestion and food	
CC	absorption disorders (Crohn's disease), neurodegenerative diseases	
CC	(Alzheimer's disease), behavioral disorders (Tourette's syndrome),	
CC	proliferative and cancerous conditions (acute myelogenous leukaemia),	
CC	allergic reactions (asthma), cardiovascular disorders (arrhythmia),	

ocular disorders (glaucoma) and infectious diseases caused by bacteria, viruses, fungi or parasites. It is also useful for screening therapeutic compounds. Serine proteases are used as immunological probes or polymorphic markers for the identification of chromosomes, cells and tissues in biological samples, identification of male contraceptive agents, delivery of compositions to targetted cells expressing a receptor for serine protease, hybridisation probes and molecular weight markers. Serine protease nucleic acids are also useful in gene therapy.

Query Match	44.7%;	Score 782;	DB 21;	Length 1222;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 782;	Conservative 0;	Mismatches 0;	Indels 0	

Qy	892	ccacaggtctgaatgcctctccacggcgatatactccctccagtggttcccactgcggact	951
Db	9	ccacaggtctgaatgcctctccacggcgatatactccctccagtggttcccactgcggact	68
Qy	952	gaggggcatgaccggcgagatcgfggagggcgctgacctcggaatgaagaatgccttg	1011
Db	69	gaggggcatgaccggcgagatcgfggagggcgctgacctcggaatgaagaatgccttg	128
Qy	1012	gcaagtgaagtgcacttcggaacacacccacatctgtgaggcagcgctcattgacgccca	1071
Db	129	gcaagtgaagtgcacttcggaacacacccacatctgtgaggcagcgctcattgacgccca	188
Qy	1072	gtgggtgctaactgccggcccaactgcttcttgcgaacccggagaaaggtctctggagggtg	1131
Db	189	gtgggtgctaactgccggcccaactgcttcttgcgaacccggagaaaggtctctggagggtg	248
Qy	1132	gaaggtgtacgcggcaccagcaacctgcacctgtgcctgaggcagcgtccattgcgga	1191
Db	249	gaaggtgtacgcggcaccagcaacctgcacctgtgcctgaggcagcgtccattgcgga	308
Qy	1192	gatcatcatcaacgagcaattacaccgatgaggagagcagctatgacatcgccctcatgcg	1251
Db	309	gatcatcatcaacgagcaattacaccgatgaggagagcagctatgacatcgccctcatgcg	368
Qy	1252	gctgtccaagcccctgacctgtgcgttcacatccacctgcttgcctcccccatgcatgg	1311
Db	369	gctgtccaagcccctgacctgtgcgttcacatccacctgcttgcctcccccatgcatgg	428
Qy	1312	acagacctttagctcctaatsgacgctgtgatcacaggctttggcaagacacaggagac	1371
Db	429	acagacctttagctcctaatsgacgctgtgatcacaggctttggcaagacacaggagac	488
Qy	1372	agatgacaagaacatcccccctctccggggaggtgcagggtcaatctatcgacttcaagaa	1431
Db	489	agatgacaagaacatcccccctctccggggaggtgcagggtcaatctatcgacttcaagaa	548
Qy	1432	atgcaatgactactgttctatgacgttaccttaccccaaggatgatgtgtcgtggga	1491
Db	549	atgcaatgactactgttctatgacgttaccttaccccaaggatgatgtgtcgtggga	608
Qy	1492	ccttcgtgggggcagagactcctgccagggagacagcggggcctcttctctgtgagca	1551
Db	609	ccttcgtgggggcagagactcctgccagggagacagcggggcctcttctctgtgagca	668
Qy	1552	gaacaaacgctgtgtacctggcaggtgtcaacagctgggggcacaggtctgtgccagagaaa	1611
Db	669	gaacaaacgctgtgtacctggcaggtgtcaacagctgggggcacaggtctgtgccagagaaa	728
Qy	1612	caaaccttgggtgttacaccaaaagtgcacagaagttcttccctggatttacacagaatgga	1671
Db	729	caaaccttgggtgttacaccaaaagtgcacagaagttcttccctggatttacacagaatgga	788
Qy	1672	ga	1673
Db	789	ga	790

RESULT 8
 AAH07577
 ID AAH07577 standard; cDNA; 852 BP.
 XX
 AC AAH07577;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:4412.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 XX 07-FEB-2001.
 PD
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 XX 29-JUL-1999; 99JP-0248036.
 PR
 PR 27-AUG-1999; 99JP-0300253.
 PR
 PR 11-YAN-2000; 2000JP-0118776.
 PR
 PR 02-MAY-2000; 2000JP-0183767.
 PR
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 PI
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 1; SEQ ID 4412; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95993 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 852 BP; 180 A; 284 C; 227 G; 153 T; 8 other;

```

Query Match      14.7%; Score 257; DB 22; Length 852;
Best Local Similarity 100.0%; Pred. No. 4.1e-101;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 ctccagaccatgagagggcagccacgggaatcatctccagcaagaacaccttcagc 60
      81 ctccagaccatgagagggcagccacgggaatcatctccagcaagaacaccttcagc 140
Db

```

QY 61 tggagcatctccagccagcagcatctccagctggagacacatccagcgggcatctccagc 120
|||||
Db 141 tggagcatctccagccagcagcatctccagctggagacacatccagcgggcatctccagc 200
|||||
QY 121 ccaggcatctccagccagcagcatctccagctggagacacatccagcgggcatctccagc 180
|||||
Db 201 ccaggcatctccagccagcagcatctccagctggagacacatccagcgggcatctccagc 260
|||||
QY 181 ccaggcatctccagctgggtacacatccagcgggcatctccagcgggcatctccagc 240
|||||
Db 261 ccaggcatctccagctgggtacacatccagcgggcatctccagcgggcatctccagc 320
|||||
QY 241 ccaggcatctccagccc 257
|||||
Db 321 ccaggcatctccagccc 337
|||||

RESULT 9
AAS91044
ID AAS91044 standard; cDNA; 1212 BP.
XX
AC AAS91044;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26848.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG26857.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 26848; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AA594564 represent novel human

CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1212 BP; 275 A; 348 C; 328 G; 261 T; 0 other;
Query Match 14.3%; Score 250; DB 23; Length 1212;
Best Local Similarity 99.7%; Pred. No. 4.2e-98;
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 804 gtgtccacccggaacacccgaggtgtgccacagggattttgccaaacagctctcaatttga 863
|||||
Db 404 gtgtccacccggaacacccgaggtgtgccacagggattttgccaaacagctctcaatttga 463
|||||
QY 864 gatacaactccaccatccaggaagcctccacaggtctgaatgcccttcccagcggtata 923
|||||
Db 464 gatacaactccaccatccaggaagcctccacaggtctgaatgcccttcccagcggtata 523
|||||
QY 924 tctccctccagtggtccaccactcgaggtgagggccatgacccgggctgtggggagggg 983
|||||
Db 524 tctctccagtggtccaccactcgaggtgagggccatgacccgggctgtggggagggg 583
|||||
QY 984 cgctggcctcgatagcaagtggtggtgcaagtgagtgacttcggcaccaccacaca 1043
|||||
Db 584 cgctggcctcgatagcaagtggtggtgcaagtgagtgacttcggcaccaccacaca 643
|||||
QY 1044 tctgtgagcgacgctcattgacgcccgagtggtgtcactgcgccactgcttcttcg 1103
|||||
Db 644 tctgtgagcgacgctcattgacgcccgagtggtgtcactgcgccactgcttcttcg 703
|||||
QY 1104 t 1104
Db 704 t 704
RESULT 10
AAL02725
ID AAL02725 standard; DNA; 926 BP.
XX
AC AAL02725;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5413.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition .
XX
PS Disclosure; SEQ ID NO 5413; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 926 BP; 189 A; 225 C; 268 G; 244 T; 0 other;

Query Match 8.9%; Score 156; DB 22; Length 926;
Best Local Similarity 100.0%; Pred. No. 1.5e-57;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1517 cagggagacagcgggggcccctctgtgtgtgacagacaacacgcgtggtacctggcaggt 1576
|||||
Db 299 cagggagacagcgggggcccctctgtgtgtgacagacaacacgcgtggtacctggcaggt 358
|||||
QY 1577 gtccacagctggggcacaggctgtgcccagagaaacaaacctgtgtgtacaccaaagt 1636
|||||
Db 359 gtccacagctggggcacaggctgtgcccagagaaacaaacctgtgtgtacaccaaagt 418
|||||
QY 1637 acagaagtcttccctggattacagcaagatggag 1672
|||||
Db 419 acagaagtcttccctggattacagcaagatggag 454
|||||
RESULT 11
AAS38533
ID AAS38533 standard; cDNA; 450 BP.
XX
AC AAS38533;
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
DE Novel human diagnostic and therapeutic gene #1591.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PS (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Rckvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
PS WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 994-995; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 450 BP; 97 A; 166 C; 112 G; 75 T; 0 other;

Query Match 8.7%; Score 152; DB 22; Length 450;
Best Local Similarity 100.0%; Pred. No. 8.5e-56;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagagaccatggagggacagccacgggaatgcattccagcagaacaccttcagc 60
|||||
Db 65 ctagagaccatggagggacagccacgggaatgcattccagcagaacaccttcagc 124
|||||
QY 61 tggagcatctccagccagcagcatctccagctggagacatccagccgggcatctccagc 120
|||||

Db 125 tggagcatctccagccagccagcattctccagctgggacacatccagccggcggcatctccagc 184
QY 121 ccaggcatctccagccagccagcattctccagctg 152
|||||
Db 185 ccaggcatctccagccagccagcattctccagctg 216
|||||
RESULT 12
AAS79452
ID AAS79452 standard; cDNA; 705 BP.
XX
AC AAS79452;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #15256.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang VT;
XX
PS WPI; 2001-639362/73.
DR P-PSDB; ABG15265.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 15256; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 705 BP; 144 A; 193 C; 210 G; 158 T; 0 other;

Query Match 8.4%; Score 146; DB 23; Length 705;
Best Local Similarity 100.0%; Pred. No. 3.2e-53;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1374 atgacaagacatcccccttctccggagggtgcaggtcactcctcagcttcaagaaat 1433
|||||
Db 218 atgacaagacatcccccttctccggagggtgcaggtcactcctcagcttcaagaaat 277
|||||
QY 1434 gcaatgactacttggtctatgacagttaccttaccctcccaaggatgctgtgctggggacc 1493
|||||
Db 278 gcaatgactacttggtctatgacagttaccttaccctcccaaggatgctgtgctggggacc 337
|||||
QY 1494 ttctgtggggcagagactcctgccag 1519
|||||
Db 338 ttctgtggggcagagactcctgccag 363
|||||

RESULT 13
ID AAL02724 standard; DNA; 926 BP.
XX
AC AAL02724;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5412.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
```

PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 XX Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX Disclosure; SEQ ID NO 5412; 1297pp + Sequence Listing; English.
 XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX Sequence 926 BP; 190 A; 224 C; 267 G; 245 T; 0 other;
 SQ
 Query Match 8.3%; Score 145; DB 22; Length 926;
 Best Local Similarity 100.0%; Pred. No. 8.3e-53;
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1376 gacagacatccccctctccggagggtgcaggtcgaatctcatcgaactcaagaatgc 1435
 Db 7 gacagacatccccctctccggagggtgcaggtcgaatctcatcgaactcaagaatgc 66
 QY 1436 atgactactgtgtatgacagttaaccttaccctcccaaggatgatgtgtgtgtgggacctt 1495
 Db 67 atgactactgtgtatgacagttaaccttaccctcccaaggatgatgtgtgtgtgggacctt 126
 QY 1496 cgtgggggagagacactctgtccagg 1520
 Db 127 cgtgggggagagacactctgtccagg 151
 RESULT 14
 AAD05802/C
 ID AAD05802 standard; DNA; 29 BP.
 XX AAD05802;
 AC
 XX
 XX 31-JUL-2001 (first entry)
 XX Human endothelias 2 gene specific 5' RACE PCR primer.
 DE
 XX

KW Human; endothelias 2; protease domain; cytostatic; vulnery; wound;
 KW neotropic; periodontitis; dermatological disorder; gene therapy; scar;
 KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
 KW chronic inflammatory disease; ocular disorder; circulatory disorder;
 KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
 KW liver cirrhosis; osteoradionecrosis, systemic sclerosis; oesophageal;
 KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
 KW systemic vasculitis; scleroderma; neoplasm; ulcer; PCR primer; RACE;
 KW rapid amplification of cDNA end; ss.
 XX
 OS Homo sapiens.
 XX WO200136604-A2.
 PN 25-MAY-2001.
 PD 17-NOV-2000; 2000WO-US31803.
 PF 18-NOV-1999; 99US-0166391.
 PR 22-SEP-2000; 2000US-0234840.
 XX (CORV-) CORVAS INT INC.
 PA Madison EL, Ong EO;
 PI WPI; 2001-336001/35.
 XX New nucleic acid encoding a protein comprising endothelias activity
 PT useful in the prevention and treatment of e.g. vascular malformations,
 PT cardiovascular disorders, and chronic inflammatory disease -
 XX Example 2; Page 117; 152pp; English.
 XX The present DNA sequence is human endothelias 2 gene specific 5' RACE
 CC (rapid amplification of cDNA ends) PCR primer which is used for obtaining
 CC a full-length cDNA clone encoding endothelias 2.
 CC The invention relates to an endothelias protein, endothelias protease
 CC domain and their corresponding nucleic acid molecules. An endothelias
 CC protein or protease domain of it is useful for the treatment and
 CC diagnosis of disorders associated with aberrant angiogenesis or undesired
 CC neovascularisation. The undesired angiogenesis is associated with
 CC disorders selected from solid neoplasm, vascular malformations and
 CC cardiovascular disorders such as angiodioma, angiolipoma,
 CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
 CC malformations, haemangiomas (Fafucci's syndrome), hereditary haemorrhagic
 CC with vascular hamartomas (Fafucci's syndrome) and Von Hippel Lindau
 CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
 CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
 CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
 CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
 CC granulations/burns, hypertrophic scars, liver cirrhosis,
 CC osteoradionecrosis, postoperative wound repairs, circulatory disorders
 CC systemic sclerosis and aberrant wound repairs, circulatory disorders
 CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
 CC dyomeolitis, sclerodactyly and teangiectasis, pyogenic granuloma and
 CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
 CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
 CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
 CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
 CC caused by ocular neovascular disease, corneal graft neovascularisation,
 CC macular degeneration, retinopathy of prematurity, retrolental
 CC fibroplasia and corneal neovascularisation. The nucleic acids of the
 CC invention are also used in gene therapy. The invention also provides
 CC method for screening compounds that modulate angiogenesis.
 XX Sequence 29 BP; 7 A; 4 C; 15 G; 3 T; 0 other;
 SQ

Query Match 1.7%; Score 29; DB 22; Length 29;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1273 gtcgctcacatccaccctgtgtgctcc 1301

Db 29 GTCCGCTCACATCCACCCCTGCTGCCTCC 1

RESULT 15
AAD05809/c
ID AAD05809 standard; DNA; 29 BP.

AC AAD05809;

XX 31-JUL-2001 (first entry)

XX Human endothelial 2-specific anti-sense RT-PCR primer.

KW Human; endothelial 2; protease domain; cytostatic; vulnery; wound;
KW nontropic; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW crest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
KW liver cirrhosis, osteoradionecrosis, systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasm; ulcer; RT-PCR primer; ss.

OS Homo sapiens.

XX WO200136604-A2.

XX 25-MAY-2001.

PF 17-NOV-2000; 2000WO-US31803.

XX 18-NOV-1999; 99US-0166391.

PR 22-SEP-2000; 2000US-0234840.

XX (CORV-) CORVAS INT INC.

PI Madison EL, Ong EO;

XX WPI; 2001-336001/35.

XX New nucleic acid encoding a protein comprising endothelial activity
PT useful in the prevention and treatment of e.g. vascular malformations,
PT cardiovascular disorders, and chronic inflammatory disease -
XX
PS Example 2; Page 120; 152pp; English.

CC The present DNA sequence is human endothelial 2-specific anti-sense RT
CC (Reverse Transcriptase)-PCR primer which is used for amplifying a 422-bp
CC fragment from the single-stranded cDNAs of human umbilical vein
CC endothelial cells (HUVEC). This 422-bp fragment spans the scavenger
CC receptor cysteine-rich domain and the serine protease domain.
CC The invention relates to an endothelial protein, endothelial protease
CC domain and their corresponding nucleic acid molecules. An endothelial
CC protein or protease domain of it is useful for the treatment and
CC diagnosis of disorders associated with aberrant angiogenesis or undesired
CC neovascularisation. The undesired angiogenesis is associated with
CC disorders selected from solid neoplasm, vascular malformations and
CC cardiovascular disorders such as angiofibroma, angiolipoma,
CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
CC malformations, haemangiomas and vascular adhesions, dyschondroplasia
CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
CC granulations-burns, hypertrophic scars, liver cirrhosis,
CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
CC systemic sclerosis and aberrant wound repairs, circulatory disorders
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
CC dyonotiloky, sclerodactyly and teangiectasis, dermatological disorders
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome

CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
CC caused by ocular neovascular disease, corneal graft neovascularisation,
CC macular degeneration, retinopathy of prematurity, retrolental
CC fibroplasia and corneal neovascularisation. The nucleic acids of the
CC invention are also used in gene therapy. The invention also provides
CC method for screening compounds that modulate angiogenesis.

XX Sequence 29 BP; 7 A; 4 C; 15 G; 3 T; 0 other;

Query Match 1.7%; Score 29; DB 22; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.013; Mismatches 0; Gaps 0;
Matches 29; Conservative 0;

QY 1273 gtccgtcacatccacccctgctgcctcc 1301
Db 29 GTCCGCTCACATCCACCCCTGCTGCCTCC 1

Search completed: August 19, 2002, 22:22:37
Job time: 9209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 17:03:57 ; Search time 291.43 Seconds
(without alignments)
10298.070 Million cell updates/sec

Title: US-09-879-792-11

Perfect score: 1748

Sequence: 1 ctcagagaccatggagaggg.....ggctgctgtgactcgagaaa 1748

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1687.4	96.5	1689	22	AA05796 Human transmembran
2	1663.4	95.2	2067	22	AA05797 Human transmembran
3	1302.8	74.5	1314	22	AAF83971 Human SER6 nucleot
c 4	1252.8	71.7	2192	22	AAH14850 Nucleotide sequenc
5	1252.8	71.7	2192	22	AAH14850 Human cDNA sequenc
c 6	1068.4	61.1	1078	22	AAF83972 Nucleotide sequenc
7	782.4	44.8	1222	21	AA02322 Human serine prote
8	535.6	30.6	852	22	AAH07577 Human cDNA clone (
9	373.2	21.4	1212	23	AA591044 DNA encoding novel

10	331.2	18.9	450	22	AA538533	Novel human diagno
11	224	12.8	705	23	AAH79452	DNA encoding novel
c 12	193.2	11.1	1854	22	AAH99574	Human protein enco
13	193.2	11.1	2038	20	AAH87154	Human protease HUP
14	193.2	11.1	2063	21	AAH37099	Human PRO1570 (UNQ
15	193.2	11.1	2063	22	AA546089	Human DNA encoding
16	193.2	11.1	2063	22	AAH92113	Human PRO1570 cDNA
17	193.2	11.1	2063	22	AAH54396	DNA encoding prote
18	193.2	11.1	2079	22	AAH13169	Human transmembran
19	191.6	11.0	2137	22	AAH13114	Human membrane-ty
20	190.8	10.9	1394	21	AA02323	Human serine prote
21	190.8	10.9	1697	21	AAH7781	Human cancer assoc
22	190.4	10.9	1305	22	AA02556	Human seripancrin
23	189.6	10.8	1479	22	AA02557	Human seripancrin
24	188.6	10.8	2106	22	AAH33397	Human colon cancer
25	188.4	10.8	1795	22	AA526880	Human cDNA encodin
26	184	10.5	2070	21	AA290471	Cancer specific ge
27	182.8	10.5	1656	22	AAH13118	Human membrane-ty
28	182.8	10.5	2412	20	AA233949	Human PRO382 nucle
29	182.8	10.5	2412	21	AAH78475	Human PRO382 (UNQ3
30	182.8	10.5	2412	22	AA545956	Human DNA encoding
31	181.2	10.4	2413	21	AAH93842	Tumour associated
32	181.2	10.4	2413	22	AA559907	Human cDNA encodin
33	177.4	10.1	1076	21	AAH12975	CDNA encoding huma
34	176.8	10.1	1738	21	AA229636	Human 20P1F12-GTC2
35	176.8	10.1	3245	21	AAH08803	Androgen-inducible
36	176.8	10.1	3245	22	AA564178	Human prostate cDN
37	176.8	10.1	3245	22	AAH93942	P1000C full length
38	176.4	10.1	1479	21	AA287786	Human tumour suppr
39	176	10.1	699	18	AAH79127	Human serine prote
40	176	10.1	1281	21	AAH97361	Human colorectal c
41	175.6	10.0	2479	21	AA290478	Ovr115 homolog pro
42	175.6	10.0	2479	21	AA287813	Human tumour suppr
43	175.6	10.0	2479	22	AA564164	Human cDNA encodin
44	175.6	10.0	2479	22	AAH13168	Human serine prote
45	175.6	10.0	2479	22	AAH93928	Human transmembran

ALIGNMENTS

RESULT 1
AA05796
ID AAD05796 standard; DNA; 1689 BP.
AC AAD05796;
XX AAD05796;
DT 31-JUL-2001 (first entry)
DE Human transmembrane serine protease (Endothelias 2-S) DNA.
KW Human; endothelias 2-S; protease domain; cytosolic; vulnary; wound;
KW angiotensin; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW chrest syndrome; atherosclerosis; haemangiomas; diabetes mellitus;
KW liver cirrhosis; osteoradionecrosis; systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;
KW transmembrane serine protease; ds.
XX Homo sapiens.
OS

Key Location/Qualifiers
CDS 1..1689
/*tag= a
/product= "Human endothelias 2-S protein"
WO200136604-A2
XX 25-MAY-2001
XX 17-NOV-2000; 2000WO-US31803.

XX 18-NOV-1999; 99US-0166391.
 PR 22-SEP-2000; 2000US-0234840.
 XX (CORV-) CORVAS INT INC.
 PA Madlison EL, Ong EO;
 XX WPI: 2001-336001/35.
 XX P-PSDB; AAE01943.
 PT New nucleic acid encoding a protein comprising endothelialase activity
 PT useful in the prevention and treatment of e.g. vascular malformations,
 PT cardiovascular disorders, and chronic inflammatory disease -
 XX Example 2; Page 135-137; 152pp; English.
 XX The present DNA sequence encodes human short form transmembrane serine
 CC protease (Endothelialase 2-S) protein.
 CC The invention relates to an endothelialase protein, endothelialase protease
 CC domain and their corresponding nucleic acid molecules. An endothelialase
 CC protein or protease domain of it is useful for the treatment and
 CC diagnosis of disorders associated with aberrant angiogenesis or undesired
 CC neovascularisation. The undesired angiogenesis is associated with
 CC disorders selected from solid neoplasm, vascular malformations and
 CC cardiovascular disorders such as angiofibroma, angiolipoma,
 CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
 CC malformations, haemangiomatosis and vascular adhesions, dyschondroplasia
 CC with vascular hamartomas (Pafucci's syndrome), hereditary haemorrhagic
 CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
 CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
 CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
 CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
 CC granulations-burns, hypertrophic scars, liver cirrhosis,
 CC osteoradionecrosis, postoperative adhesion, pyogenic granuloma and
 CC systemic sclerosis and aberrant wound repairs, circulatory disorders
 CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
 CC dyomyolysis, sclerodactyly and teangiectasis, dermatological disorders
 CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
 CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
 CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
 CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
 CC caused by ocular neovascular disease, corneal graft neovascularisation,
 CC macular degeneration, retinopathy of prematurity, retrolental
 CC fibroplasia and corneal neovascularisation. The nucleic acids of the
 CC invention are also used in gene therapy. The invention also provides
 CC method for screening compounds that modulate angiogenesis.
 XX Sequence 1689 BP; 367 A; 529 C; 470 G; 323 T; 0 other;

Query Match 96.5%; Score 1687.4; DB 22; Length 1689;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 11 atggagaggagcagccagcgggaatgcattctccagcaagaacaccttcagctggagcatct 70
 Db 1 atggagaggagcagccagcgggaatgcattctccagcaagaacaccttcagctggagcatct 60
 Qy 71 ccagccaggcatctccagctgggacacctccagccggggcatctccagccaggcatct 130
 Db 61 ccagccaggcatctccagctgggacacctccagccggggcatctccagccaggcatct 120
 Qy 131 ccagccaggcatctccagctgggacacctccagccggggcatctccagccaggcatct 190
 Db 121 ccagccaggcatctccagctgggacacctccagccggggcatctccagccaggcatct 180
 Qy 191 ccagctgggtacacctccagccggggcatctccagccggggcatctccagccaggcatct 250
 Db 181 ccagctgggtacacctccagccggggcatctccagccggggcatctccagccaggcatct 240
 Qy 251 ccagccggggcatctccggctctggatcactttccaggctctccatccggcaggtcatca 310
 Db 1321 gagacctgctggatccacaggctttggcaagaccgggagagacagatgacagacatcccc 1380

QY 1391 ttctccgggaggtgcaggtcaatctcatcgaacttcaagaaatgcaatgactacttggtc 1450
|||||
Db 1381 ttctccgggaggtgcaggtcaatctcatcgaacttcaagaaatgcaatgactacttggtc 1440
QY 1451 tatgacagttacattaccccaagagatgctgtctggtgggacattcgtggtggcagagac 1510
|||||
Db 1441 tatgacagttacattaccccaagagatgctgtctggtgggacattcgtggtggcagagac 1500
QY 1511 tctgcaggagacagcgggggaccttctgtctgtgagcagaaacccctggtacctg 1570
|||||
Db 1501 tctgcaggagacagcgggggaccttctgtctgtgagcagaaacccctggtacctg 1560
QY 1571 gcaggtgtccacagctgggacagcgtgtggccagagaaacaaacctgtgtgtacacc 1630
|||||
Db 1561 gcaggtgtccacagctgggacagcgtgtggccagagaaacaaacctgtgtgtacacc 1620
QY 1631 aaagtgcagaaagtcttccctggatttacagcaagatggagcaggtgcgattcaga 1690
|||||
Db 1621 aaagtgcagaaagtcttccctggatttacagcaagatggagcaggtgcgattcaga 1680
QY 1691 aaatcctaa 1699
|||||
Db 1681 aaatcctaa 1689

RESULT 2
AAD05797
ID AAD05797 standard; DNA; 2067 BP.
XX
AC AAD05797;
XX
DT 31-JUL-2001 (first entry)
XX
DE Human transmembrane serine protease (Endotheliase 2-L) DNA.
XX
KW Human; endotheliase 2-L; protease domain; cytotstatic; vulneryary; wound;
KW nontropic; periodontitis; dermatological disorder; gene therapy; scar;
KW angiogenesis; cardiovascular disorder; psoriasis; neovascular disease;
KW chronic inflammatory disease; ocular disorder; circulatory disorder;
KW crest syndrome; atherosclerosis; haemangiomatosis; diabetes mellitus;
KW liver cirrhosis; osteoradionecrosis; systemic sclerosis; oesophageal;
KW inflammatory bowel disease; fracture; rheumatoid arthritis; retinopathy;
KW systemic vasculitis; scleroderma; neoplasm; ulcer; burn;
KW transmembrane serine protease; ds.
XX
OS Homo sapiens.

FH Key Location/Qualifiers
FT CDS 1..2067
FT FT /*tag= a
FT FT /product= "Human endotheliase 2-L protein"
XX
PN WO200136604-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31803.
XX
PF 18-NOV-1999; .99US-0166391.
PR 22-SEP-2000; 2000US-0234840.
XX
XX (CORV-) CORVAS INT INC.
XX
PI Madison EL, Ong EO;
XX
XX WPI; 2001-336001/35.
DR P-PSDB; AAE01944.
XX
XX New nucleic acid encoding a protein comprising endotheliase activity
PT useful in the prevention and treatment of e.g. vascular malformations,
PT cardiovascular disorders, and chronic inflammatory disease -
XX

Example 2; Page 139-142; 152pp; English.

PS The present DNA sequence encodes human long form transmembrane serine
XX protease (Endotheliase 2-L) protein.
CC The invention relates to an endotheliase protein, endotheliase protease
CC domain and their corresponding nucleic acid molecules. An endotheliase
CC protein or protease domain of it is useful for the treatment and
CC diagnosis of disorders associated with aberrant angiogenesis or undesired
CC neovascularisation. The undesired angiogenesis is associated with
CC disorders selected from solid neoplasm, vascular malformations and
CC cardiovascular disorders such as angiofibroma, angiolipoma,
CC atherosclerosis, restenosis/reperfusion injury, arteriovenous
CC malformations, haemangiomatosis and vascular adhesions, dyschondroplasia
CC with vascular hamartomas (Fafucci's syndrome), hereditary haemorrhagic
CC telangiectasia (Rendu-Osler-Weber syndrome) and Von Hippel Lindau
CC syndrome, chronic inflammatory diseases such as diabetes mellitus,
CC haemophilic joints, inflammatory bowel disease, nonhealing fractures,
CC periodontitis, psoriasis, rheumatoid arthritis, venous stasis ulcers,
CC osteoradionecrosis, hypertrophic scars, liver cirrhosis,
CC systemic sclerosis and aberrant wound repairs, circulatory disorders
CC Raynaud's phenomenon, crest syndromes such as calcinosis, oesophageal,
CC dyomotility, sclerodactyly and teangiectasis, dermatological disorders
CC such as systemic vasculitis, scleroderma, pyoderma gangrenosum,
CC vasculopathy, venous, arterial ulcers, Sturge-Weber syndrome, Port-wine
CC stains, blue rubber bleb nevus syndrome, Klippel-Trenaunay-Weber syndrome
CC and Osler-Weber-Rendu syndrome and ocular disorders such as blindness
CC caused by ocular neovascular disease, corneal graft neovascularisation,
CC macular degeneration, retinopathy of prematurity, retrolental
CC fibroplasia and corneal neovascularisation. The nucleic acids of the
CC invention are also used in gene therapy. The invention also provides
CC method for screening compounds that modulate angiogenesis.
XX
SQ Sequence 2067 BP; 463 A; 604 C; 540 G; 460 T; 0 other;

Query Match 95.2%; Score 1663.4; DB 22; Length 2067;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 atggagaggagacacgacgggaatgcatctccagcaagaacacattcagctggagcatct 70
|||||
Db 1 atggagaggagacacgacgggaatgcatctccagcaagaacacattcagctggagcatct 60
QY 71 ccagccaggcatctccagctgggacacctccagcggcgcatctccagccaggcatct 130
|||||
Db 61 ccagccaggcatctccagctgggacacctccagcggcgcatctccagccaggcatct 120
QY 131 ccagccaggcatctccagctgggacacctccgggcgcgcatctccagccaggcatct 190
|||||
Db 121 ccagccaggcatctccagctgggacacctccgggcgcgcatctccagccaggcatct 180
QY 191 ccagctggttacacctccagcggcgcatctccagcggcgcgcatctccagccaggcatct 250
|||||
Db 181 ccagctggttacacctccagcggcgcatctccagcggcgcgcatctccagccaggcatct 240
QY 251 ccagccggcgcatctccgctctggcatacttccaggtccctccatccgagagcatca 310
|||||
Db 241 ccagccggcgcatctccgctctggcatacttccaggtccctccatccgagagcatca 300
QY 311 tccgcccagggtcagcctcgtgtgacaacctcccacacagagtgtagctttagagaaca 370
|||||
Db 301 tccgcccagggtcagcctcgtgtgacaacctcccacacagagtgtagctttagagaaca 360
QY 371 ccagtggtgggctgtacctccatccatctctccaggtccagcagcagcagggcc 430
|||||
Db 361 ccagtggtgggctgtacctccatccatctctccaggtccagcagcagcagggcc 420
QY 431 accagagagagccaggtacgacctgcccaggttccacctggcgagggcgagagcag 490
|||||
Db 421 accagagagagccaggtacgacctgcccaggttccacctggcgagggcgagagcag 480
QY 491 ctaccgctcagcgggtgctgctcctcctcctcctcctggtggttctgcctcctc 550

Db 481 ctaccgctcaggtgctgctccctccctcatgctccctggtgttctcgtcatcctc 540
QY 551 ttcaggttctggcagggccacacagagatcaggtacaagagcagagggagagctgcc 610
Db 541 ttcaggttctggcagggccacacagagatcaggtacaagagcagagggagagctgcc 600
QY 611 aagcacgctgttgcgtgtgacgggtggtggaactgcaagctgaagatgacgagctgggc 670
Db 601 aagcacgctgttgcgtgtgacgggtggtggaactgcaagctgaagatgacgagctgggc 660
QY 671 tgcgtgaggttgcagtggaagctctgctgttaaaatctactctggtccctcccatcag 730
Db 661 tgcgtgaggttgcagtggaagctctgctgttaaaatctactctggtccctcccatcag 720
QY 731 tggcttcccatctgtagcgaactggaatgactcctactcagagaaacctgcccagcag 790
Db 721 tggcttcccatctgtagcgaactggaatgactcctactcagagaaacctgcccagcag 780
QY 791 ctgggtttcagagtgctcaccggacaacccaggtgtgcccacagggattttgccaacgc 850
Db 781 ctgggtttcagagtgctcaccggacaacccaggtgtgcccacagggattttgccaacgc 840
QY 851 ttctcaatttgagatacaactccaccatccaggaaagcctccacaggtctgaatgccct 910
Db 841 ttctcaatttgagatacaactccaccatccaggaaagcctccacaggtctgaatgccct 900
QY 911 tcccagcgttatctccctcagtggttcccactcgggactcggagccatgaccggggg 970
Db 901 tcccagcgttatctccctcagtggttcccactcgggactcggagccatgaccggggg 960
QY 971 atcgttggagggcgctggcctcgatagcaagtgagccttggaagtgaagtgcacttc 1030
Db 961 atcgttggagggcgctggcctcgatagcaagtgagccttggaagtgaagtgcacttc 1020
QY 1031 ggcacacacacacatctgtgagggcgcctcattgacgccagtggtgctcactccgcc 1090
Db 1021 ggcacacacacacatctgtgagggcgcctcattgacgccagtggtgctcactccgcc 1080
QY 1091 cactgcttcttgacccggaggaaggtctctggagggtggaagtgtacgcgggcacc 1150
Db 1081 cactgcttcttgacccggaggaaggtctctggagggtggaagtgtacgcgggcacc 1140
QY 1151 agcaacctgcaccagtgctccctgagggcagcctccattgacggagatcatcaacaagaat 1210
Db 1141 agcaacctgcaccagtgctccctgagggcagcctccattgacggagatcatcaacaagaat 1200
QY 1211 tacaccgatgaggagacgactatgacatcgccctcatcggtgtccaaagccctgacc 1270
Db 1201 tacaccgatgaggagacgactatgacatcgccctcatcggtgtccaaagccctgacc 1260
QY 1271 ctgtccgtcacatccaccctgcttgcctcccatgcgatgacagacctttagcctcaat 1330
Db 1261 ctgtccgtcacatccaccctgcttgcctcccatgcgatgacagacctttagcctcaat 1320
QY 1331 gagacctgtgatacacaggtcttggcaagaccagggagacagatgacaagaatcccc 1390
Db 1321 gagacctgtgatacacaggtcttggcaagaccagggagacagatgacaagaatcccc 1380
QY 1391 ttctccggggaggtgcaggttcaatctcatcgacttcaagaatgcaatgactacttgctc 1450
Db 1381 ttctccggggaggtgcaggttcaatctcatcgacttcaagaatgcaatgactacttgctc 1440
QY 1451 tatgacagttaccctaccaccaagatgatgtgtctggggaccctctgtggggcagagac 1510
Db 1441 tatgacagttaccctaccaccaagatgatgtgtctggggaccctctgtggggcagagac 1500
QY 1511 tcctgcaggagagacggggggcctcttctgtgtgagcagaacaaccctgtactctg 1570
Db 1501 tcctgcaggagagacggggggcctcttctgtgtgagcagaacaaccctgtactctg 1560
QY 1571 gcaggtgtcaccagctggggccacaggtgtgtggccagagaacaaccctgtgtgtacac 1630

Db 1561 gcaggtgtcaccagctggggccacaggtgtgtgccagagaaacacacctgtgtgtacac 1620
QY 1631 aaagtacagaagttctccctggatttacagcaagatggagac 1675
Db 1621 aaagtacagaagttctccctggatttacagcaagatggagac 1665
RESULT 3
AAF83971
ID AAF83971 standard; DNA; 1314 BP.
XX AC AAF83971;
XX AC AAF83971;
DT 06-AUG-2001 (first entry)
XX Human SER6 nucleotide sequence.
XX Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;
KW cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.
KW immunosuppressive; antiasthmatic; antifibrinolytic; SER6; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 1..1266
FT /*tag= a
FT /product= "SER6"
XX WO200136645-A2.
PD 25-MAY-2001.
XX PF 17-NOV-2000; 2000WO-US31744.
XX PR 17-NOV-1999; 99US-0165986.
PR 09-FEB-2000; 2000US-0181347.
PR 03-APR-2000; 2000US-0194195.
PR 05-APR-2000; 2000US-0194839.
PR 07-APR-2000; 2000US-0195637.
PR 13-APR-2000; 2000US-0197080.
PR 15-SEP-2000; 2000US-0232677.
PR 16-NOV-2000; 2000US-0165986.
XX (CURA-) CURAGEN CORP.
FA Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Taupier R, Rastelli L, Herrmann JL;
XX WPI: 2001-336006/35.
DR P-PSDB; AAB85042.
XX New isolated SERX polynucleotides and polypeptides related to the
PT serine/threonine kinase family of proteins, useful for determining and
PT treating diseases related to altered levels of the polypeptide -
XX Claim 9; Page 33; 127pp; English.
XX The invention relates to serine/threonine kinase-like proteins, termed
CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
CC via a recombinant expression vector in a host cell in gene therapy
CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
CC polypeptides, polynucleotides, antibodies can also be used to treat or
CC prevent a pathology associated with SERX. Disorders associated with
CC aberrant SERX expression or activity include breast or testicular cancer,
CC cardiovascular defect, autoimmune disorders and asthma. Disorders
CC associated with aberrant SERX-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
XX represents the nucleotide sequence of human SER6.
SQ Sequence 1314 BP; 288 A; 384 C; 376 G; 266 T; 0 other;

Query Match			
Best Local Similarity 74.5%; Score 1302.8; DB 22; Length 1314;			
Matches 1304; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	434	aggagagccagtagcagcctcccaagttccacttgcggggagggcccaagaacagcta	493
Db	1	atggagagccagtagcagcctcccaagttccacttgcggggagggcccaagaacagcta	60
QY	494	ccgctcatcgggtgcgtgctcctcctcattgccttgcgttgcgtcctcctcttc	553
Db	61	ccgctcatcgggtgcgtgctcctcctcattgccttgcgttgcgtcctcctcttc	120
QY	554	cagttctggcaggccacacagggatcagggtacagagagcagagggagagctgccaaag	613
Db	121	cagttctggcaggccacacagggatcagggtacagagagcagagggagagctgccaaag	180
QY	614	cacgctctgcgtgacgggtggtgactgcaagctgaagctgaagagtagcagagctggctgc	673
Db	181	caecgtctgcgtgacgggtggtgactgcaagctgaagagtagcagagctggctgc	240
QY	674	gtgaggttgcagtgacagagctctcgtcttaaaatctactctggtctcccatcagtg	733
Db	241	gtgaggttgcagtgacagagctctcgtcttaaaatctactctggtctcccatcagtg	300
QY	734	cttcccatctgtagcagcaactggaatgactcctactcagagaagacgtgccagcgtg	793
Db	301	cttcccatctgtagcagcaactggaatgactcctactcagagaagacgtgccagcgtg	360
QY	794	ggttccagagtgctcaccgacacacagaggttgcacacagggatttgcacacagcttc	853
Db	361	ggttccagagtgctcaccgacacacagaggttgcacacagggatttgcacacagcttc	420
QY	854	tcaatcttgagatacaactccaccatccaggaagcctccacaggtctgaatgcccttc	913
Db	421	tcaatcttgagatacaactccaccatccaggaagcctccacaggtctgaatgcccttc	480
QY	914	cagcgttatctcctccagtgctccactgagagctgagggccatgacggggcgatc	973
Db	481	cagcgttatctcctccagtgctccactgagagctgagggccatgacggggcgatc	540
QY	974	gtgggagggcgctggcctcgatagcaagtgccttggcagtgagctgcacttcggc	1033
Db	541	gtgggagggcgctggcctcgatagcaagtgccttggcagtgagctgcacttcggc	600
QY	1034	accaccacatctgtgaggaacagctcattgacgccagtggtgctcactgcgcgccac	1093
Db	601	accaccacatctgtgaggaacagctcattgacgccagtggtgctcactgcgcgccac	660
QY	1094	tgtcttctgtagccgggagaggtcctgagggctggaagtgtagcgggccaccagc	1153
Db	661	tgtcttctgtagccgggagaggtcctgagggctggaagtgtagcgggccaccagc	720
QY	1154	aactgcaccagtgctgcagcagctccattgcccagatcatcaacagcaattac	1213
Db	721	aactgcaccagtgctgcagcagctccattgcccagatcatcaacagcaattac	780
QY	1214	accgatgaggaacgactatgacatgcctcctatgcgctgtccaaagccctgacccgtg	1273
Db	781	accgatgaggaacgactatgacatgcctcctatgcgctgtccaaagccctgacccgtg	840
QY	1274	tcgctcaatccacccctgcttgcctcccatgcatggagacacctttagcctcaatgag	1333
Db	841	tcgctcaatccacccctgcttgcctcccatgcatggagacacctttagcctcaatgag	900
QY	1334	acctgctggatcacaggcttggcaagaccagagagacagatgacaagatcccccttc	1393
Db	901	acctgctggatcacaggcttggcaagaccagagagacagatgacaagatcccccttc	960
QY	1394	ctccggaggtgcaggtccaatctcatcagcttcaagaaatgcaatgactacttggtctat	1453
Db	961	ctccggaggtgcaggtccaatctcatcagcttcaagaaatgcaatgactacttggtctat	1020
QY	1454	gacagttaccttaccaccaaggtatgtgtgctggggaccttgcgtggggcagagactcc	1513

Db	1021	gacagttacattaccaccaaggtatgtgtgctgggaccttgcgtggggcagagactcc	1080
QY	1514	tgcaggagagacagcgggggcccctcttctgtgtgagcagaacaaccgctgttacctggca	1573
Db	1081	tgcaggagagacagcgggggcccctcttctgtgtgagcagaacaaccgctgttacctggca	1140
QY	1574	ggtgtcaccagctggggccacaggtgtgtggccagagaaacacccctggtgtgtacaccaa	1633
Db	1141	ggtgtcaccagctggggccacaggtgtgtggccagagaaacacccctggtgtgtacaccaa	1200
QY	1634	gtgacagaagtcttccctggtatttacagcaagatggagagcggaggtgcgatttcagaaaa	1693
Db	1201	gtgacagaagtcttccctggtatttacagcaagatggagagcggaggtgcgatttcacaaaa	1260
QY	1694	tcctaacacagctggcctgctctctgcacagcacccggtgctgtga	1739
Db	1261	tcctaacacagctggcctgctctctgcacagcacccggtgctgtga	1306
RESULT 4			
AAF83973/C			
ID	AAF83973 standard; DNA; 1314 BP.		
XX	AAF83973;		
XX	06-AUG-2001 (first entry)		
DE	Nucleotide sequence of SER8, a partial reverse complement of SER6.		
XX	Serine/threonine kinase-like protein; SERX; breast; testicular; cancer;		
KW	cardiovascular; autoimmune disorder; cytostatic; cardiant; gene therapy.		
KW	immunosuppressive; antiasthmatic; antifibrinolytic; SER6; SER8; ss.		
OS	Homo sapiens.		
XX	WO200136645-A2.		
XX	25-MAY-2001.		
XX	17-NOV-2000; 2000WO-US31744.		
XX	17-NOV-1999; 99US-0165986.		
PR	09-FEB-2000; 2000US-0181347.		
PR	03-APR-2000; 2000US-0194195.		
PR	05-APR-2000; 2000US-0194839.		
PR	07-APR-2000; 2000US-0195637.		
PR	13-APR-2000; 2000US-0197080.		
PR	15-SEP-2000; 2000US-0232677.		
PR	16-NOV-2000; 2000US-0165986.		
PA	(CURA-) CURAGEN CORP.		
XX	Quinn KE, Spytek KA, Mejumder K, Vernet C, Burgess C, Fernandes E;		
PI	Traupier R, Rastelli L, Herrmann JL;		
XX	WPI; 2001-336006/35.		
XX	New isolated SERX polynucleotides and polypeptides related to the		
PT	serine/threonine kinase family of proteins, useful for determining and		
PT	treating diseases related to altered levels of the polypeptide -		
XX	Claim 9; Page 38; 127pp; English.		
XX	The invention relates to serine/threonine kinase-like proteins, termed		
CC	SERX. The SERX polynucleotides can be used to express SERX protein (e.g.		
CC	via a recombinant expression vector in a host cell in gene therapy		
CC	applications), to detect SERX mRNA (e.g. in a biological sample) or a		
CC	genetic lesion in a SERX gene and modulate SERX activity. The SERX		
CC	polypeptides, polynucleotides, antibodies can also be used to treat or		
CC	prevent a pathology associated with SERX. Disorders associated with		
CC	aberrant SERX expression or activity include breast or testicular cancer,		
CC	cardiovascular defect, autoimmune disorders and asthma. Disorders		

CC associated with aberrant SER5-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
CC represents the nucleotide sequence of human SER8, a partial reverse
CC complement of SER6.
XX
SQ Sequence 1314 BP; 266 A; 376 C; 384 G; 288 T; 0 other;

Query Match 74.5%; Score 1302.8; DB 22; Length 1314;
Best Local Similarity 99.8%; Pred. No. 4.3e-256;
Matches 1304; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 434 agggagagccagtgacgagctgcccaagttcacctgcccgggagggccagagcagcta 493
DB 1314 ATGACAGCCAGGTACGAGCTGCCCAAGTTCACCTGGCGGAGGCCAGAGCAGCTA 1255
QY 494 cgcctcatcgggtgcgtctctctctctctcattgccttggtgttgcctcatcctctctc 553
DB 1254 CCGCTCATCGGGTGGCTGCTCCTCTCATTTGCCCTGGTTCGCTCATCTCTCTTC 1195
QY 554 cagttctgcagggccacacagggatcaggtatcaagagagcagaggagagctgtcccaag 613
DB 1194 CAGTTCTGCGAGGGCCACACAGGGATCAGGTACAAGGACAGAGGAGAGCTGTCCCAAG 1135
QY 614 cagcgtgttcgctgtgacggggtggtggactgcaagctgaagtgacgagctgagctgc 673
DB 1134 CAGCGTGTTCGCTGTGACGGGGTGGTGACTGCAAGCTGAAGAGTGACGAGTGGGCTGC 1075
QY 674 gtgaggttgactgggacagtcctctgtcttaaaatctactctgggtctcccatcagtg 733
DB 1074 GTGAGGTTTACTGGGACAAAGTCTCTGCTTAAATCTACTCTGGGTCTCCCATCAGTGG 1015
QY 734 ctcccatctgacagcaactggaatgactctactcaagaaagacacctgcccagagctg 793
DB 1014 CTTCCTATCTGACAGCAACTGGAATGACTCCTACTCAGAGAAGACCTGCCAGCAGCTG 955
QY 794 ggttcgagagtgtccaccggacaaccagggttgcccacagggattttgccaacagcttc 853
DB 954 GGTTCGAGAGTGTCCACGGACAACCGAGGTGGCCACAGGGATTTGCCAACAGCTTC 895
QY 854 tcaattctgagatacaactccaccatccacggaaacccctccacaggtctgaatgccttcc 913
DB 894 TCAATCTTGAGATACAACTCCACCATCCAGGAAGACCTCCACAGGCTGTAATGCCCTTCC 835
QY 914 cagcgttatctcctccagtgctccactgcgagctgagggccatgacggggcgagtc 973
DB 834 CAGCGGTATATCTCCTCTCAGTGTTCCTCCTCGGACTGAGGGGCCATACCGGGCGGATC 775
QY 974 gtgggagggcgctggcctcgatagcaagtggccttgccagtgtgctgcaacttcggc 1033
DB 774 GTGGAGGGCGCTGGCCTCGGATAGCAAGTGGCTTGGCAAGTGAGTCTGCACCTTCGGC 715
QY 1034 accaccacatctgtggaggcagctcattgacgcccagtggtgctcaactgcgcgccac 1093
DB 714 ACCACCCACATCTGTGGAGGACGCTCATTTACGCCACAGTGGTGTCTCACTGCCGCCAC 655
QY 1094 tgcctctcgagccggagaaaggtcctgagggctggaaggtgtaagggcgagggcaccacg 1153
DB 654 TGCTTCTTGTCGTCGCCGGAGAGGTCTGGAGGGCTGGAAGGTGTACCGGGGCACCCAGC 595
QY 1154 aacctgcaccagttgctgagggcgcctccattgcccagatcatcatcaacagcaattac 1213
DB 594 AACCTGCACCAAGTTGCTTGAGGAGCCTCCATTGCCGAGATCATCATCAACAGCAATTAC 535
QY 1214 accgatgaggagagcagactatgacatgcgcctcatgcccgtgttccaaagcccctgaccc 1273
DB 534 ACCGATGAGGAGGACGACTATGATATGCCCTCATATGGGGCTGTCCAAGCCCTGACCCCTG 475
QY 1274 tccgctcacatccacctgcttgcctcccatgcatgacagacacctttagctcaatgag 1333
DB 474 TCCGCTTACATCCACCTCTGCTTCCCATGTCATGACAGAGCTTTAGCTCTCAATGAG 415
QY 1334 acctgctggatcacaggctttgttgcaagaccagggagagacagatgacaagacatcccttc 1393

DB 414 ACCTGCTGATCACAGGCTTTGGCAAGACCAGGAGACAGATGACAAGACATCCCCCTTC 355
QY 1394 ctccggaggtgcaggtcgaatctcatctcaagaataatgcaatgactacttggtctat 1453
DB 354 CTCCGGGAGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTACTTGGTCTAT 295
QY 1454 gacagttacttaccaccaaggatgagtgtgctgctgggacctgctggtgggagagactcc 1513
DB 294 GACAGTTAGCTTTACCCCAAGGATGATGTGTGCTGGGAGCTTCGTGGGGCAGAGACTCC 235
QY 1514 tcccagggagacagcgggggacctctgtgtgagcagaacaacacgcctgggtacctggca 1573
DB 234 TCCAGGAGAGACAGCGGGGGGCTCTTGTGTGTGAGCAGAAACACCGCTGGTACCTGGCA 175
QY 1574 ggtgtaccagctggggcagcaggtgtggtccagagaaacaaacctggtgtgtacaccaa 1633
DB 174 GGTGTACACAGCTGGGGCACAGCTGTGGCCAGAGAAACAACTGCTGTGTACACAAA 115
QY 1634 gtgacagaagttcttccctggatttacagaagatggagcaggtgcgattcagaaaa 1693
DB 114 GTGACAGAAGTTCTTCCCTGGATTTACAGCAAGATGGAGAGCGAGTGGATTACAAAA 55
QY 1694 tccataccagctggcctgctgctgtgacagcaccggctgctgtga 1739
DB 54 TCCTAACAGCTGGGCTGCTGCTCTGCTGACAGACCGGCTGCTGTGA 9

RESULT 5

AAH14850
ID AAH14850 standard; cDNA; 2192 BP.
XX
AC AAH14850;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12681.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.

XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12681; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the

XX Quinn KE, Spytek KA, Majumder K, Vernet C, Burgess C, Fernandes E;
PI Taupier R, Rastelli L, Herrmann JL;
XX WPI; 2001-336006/35.
XX New isolated SERX polynucleotides and polypeptides related to the
PT serine/threonine kinase family of proteins, useful for determining and
PT treating diseases related to altered levels of the polypeptide -
XX Claim 9; Page 37; 127pp; English.
XX The invention relates to serine/threonine kinase-like proteins, termed
CC SERX. The SERX polynucleotides can be used to express SERX protein (e.g.
CC via a recombinant expression vector in a host cell in gene therapy
CC applications), to detect SERX mRNA (e.g. in a biological sample) or a
CC genetic lesion in a SERX gene and modulate SERX activity. The SERX
CC polypeptides, polynucleotides, antibodies can also be used to treat or
CC prevent a pathology associated with SERX. Disorders associated with
CC aberrant SERX expression or activity include breast or testicular cancer,
CC cardiovascular defect, autoimmune disorders and asthma. Disorders
CC associated with aberrant SERX-8 expression include, blood clotting
CC hematopoietic and tumor related disorders. The present sequence
CC represents the nucleotide sequence of human SER7, a partial reverse
CC complement of SER6.
XX Sequence 1078 BP; 221 A; 296 C; 319 G; 242 T; 0 other;
SQ

Query Match 61.1%; Score 1068.4; DB 22; Length 1078;
Best Local Similarity 99.9%; Pred. No. 2.1e-208;
Matches 1069; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 670 ctgcgtgaggttgactgggacaaagtctctgttaaaatctactctggtgctccccaatca 729
Db 1078 CTGCGTGAGGTTGACTGGGACAACTCTGCTTAAATCTACTCTGGGCTCCCATCA 1019

QY 730 gtggcttccactctctagcagcaactggaatgaactcctactcactagagaactcgccagca 789
Db 1018 GTGGCTTCCCACTCTAGCAGCAACTGGAATGACTCTCTACTCAGAGAACCTCGCAGCA 959

QY 790 gctgggttccagagtgctccacgacacacgagagtggttccacagaggtatttgcacacag 849
Db 958 GTGGGTTTCGAGAGTGTCTACCGACACACCGAGAGTTGCCACACAGGATTTGCCAACAG 899

QY 850 cttctcaatcttgatagatacaactccaccatccaggaagcctccacaggtctggaatgcc 909
Db 898 CTTCTCAATCTTGAGATACAACTCCACCATCCAGAAAGCCTCCACAGGTCTGAAATGCC 839

QY 910 ttccagcggatatctccctccagtggttccactcggagactgagggccatgaccggcg 969
Db 838 TTCCACAGCGGTATATCTCCCTCCAGTGTCCCACTGCGGACTGAGGGCCATGACCGGCG 779

QY 970 gatcgtggagggcgctggctcgatagcaagtggccttggcaagtggatcgtgcaatt 1029
Db 778 GATCTGGGAGGGGCGCTGGCTCGGATAGCAAGTGGCTTGGCAAGTGAAGTCTGCATTT 719

QY 1030 cggcaccacccacatctgtgaggcagcactcattgacgccagtggtggtcactgcgcg 1089
Db 718 CGGCACACCCACATCTGTGGAGGACGCTATTCACGCCAGTGCGGTGCTCACTGCCGC 659

QY 1090 ccaactgcttcttgacccgggagaaagctcgtgagggctggaaggtgacggggcgac 1149
Db 658 CCACATGCTTCTTCGTGACCCCGGAGAAAGGTCTTGAGGGGTGGAAGGTGTACGGCGGAC 599

QY 1150 cagcaacctgcacagttgctgagcagcctccattcccgagatcatcatcaacagcaa 1209
Db 598 CAGCAACCTGCACCAAGTTGCTCTGAGCAGCCTCCATTGCCAGATCATCATCAACAGAA 539

QY 1210 ttacaccgatgaggaggacgactatgacatcgccctcactcggtgtccaaagccctgac 1269
Db 538 TTACACCGATGAGGAGGACGACTATGACATCGCCCTCATCGGCTGTCCAAAGCCCTGAC 479

QY 1270 cctgtccgtcacatccaccctctgttgcctccccatgcatggagacgaacttagcctcaa 1329
Db 478 CCTGTCCGCTCATATCCACCCCTGTCTGCTCCCATGATGGACAGACTTTAGCTCAA 419

QY 1330 ttagacctgtgatcacagagcttggcaagaccaggagagacagatgacaaacatcccc 1389
Db 418 TGAGACCTGCTGGATCACAGGCTTTGGCAAGACCAGGAGACAGATGACAAACATCCCC 359

QY 1390 cttctccggaggtgaggtcaactctcatcgacttcaagaaatcaatgactacttgg 1449
Db 358 CTTCTCCGGGAGGTGCAGGTCAATCTCATTCACCTTCAGAAATGCAATGACTACTTGGT 299

QY 1450 ctatgacagttaccttaccctcccaaggtatgtgtctggggacctctgtgggggagaga 1509
Db 298 CTATGACAGTTACCTTACCCCAAGGATGATGTGTCTGGGACCTTCGTGGGGCAGAGA 239

QY 1510 ctctgtccaggagagacaggggggctctgtctgtgagcagaacaaccgtgtacct 1569
Db 238 CTCTCTCCAGGAGACAGCGGGGGCCCTCTGTCTGTGAGCAGAACACCCGCTGTACT 179

QY 1570 ggcaggtgtcacacagctgggacaggtgtggtgagcagaacaacacctggtgtgtacac 1629
Db 178 GGCAGGTGTCAACAGCTGGGGCAGAGGCTGTGGCCAGAGAACAAACCTGGTGTGTACAC 119

QY 1630 caaagtgcagaagtcttccctcgtgatttacagcaagatgagagcgaggtgcgtacag 1689
Db 118 CAAAGTGACAGAACTTCTTCCCTGGATTTACAGCAAGATGGAGAGCGAGGTGCGATTAC 59

QY 1690 aaatcctaacagctggcctgctctgcacagcaccgactgtga 1739
Db 58 AAATCTTAACAGCTGGCTGCTCTCTGCACAGCACCGGCTGCTGTGA 9

RESULT 7

AAD02322
ID AAD02322 standard; cDNA; 1222 BP.

XX AAD02322;

DT 28-MAR-2001 (first entry)

XX Human serine protease #3, encoding HATEE38 cDNA clone.

XX Human; serine protease; osteopathic; immunosuppressive; antiallergic;
KW antiinflammatory; cytostatic; cardiac; neuroprotective; nootropic;
KW neuroleptic; vulnary; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 76..1194

FT /*tag= a
FT /product= "Human serine protease #3 from clone HATEE38"
CDS 76..963
FT /*tag= b
FT /product= "Human serine protease #3 fragment from
FT clone HATEE38"
FT /note= "Does not include stop codon"
FT /partial

PN WO200068247-A2.

XX 16-NOV-2000.

XX 05-MAY-2000; 2000WO-US12207.

CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 852 BP; 180 A; 284 C; 227 G; 153 T; 8 other;

Query Match 30.6%; Score 535.6; DB 22; Length 852;
Best Local Similarity 93.1%; Pred. No. 5.9e-100;
Matches 632; Conservative 0; Mismatches 22; Indels 25; Gaps 6;

QY 1 ctcagagaccatgagaggacacccagcggaatgcattctccagcaagaacaccttcagc 60
DB 81 ctcagagaccatgagaggagggagacccagcggaatgcattctccagcaagaacaccttcagc 140
QY 61 tggagcatctccagccagggcatctccagctgggacacctccagcgggcatctccagc 120
DB 141 tggagcatctccagccagggcatctccagctgggacacctccagcgggcatctccagc 200
QY 121 ccaggcatctccagccagggcatctccagctgggacacctccagcgggcatctccagc 180
DB 201 ccaggcatctccagccagggcatctccagctgggacacctccagcgggcatctccagc 260
QY 181 ccaggcatctccagctggatcacctccagcgccgggcatctccagcgccg----- 228
DB 261 ccaggcatctccagctggatcacctccagcgccgggcatctccagcgccgggcatctccagc 320
QY 229 ---ggcatctccagccagggcatctccagccgggcatctccagctgggcatctccagc 285
DB 321 ccaggcatctccagccagggcatctccagccgggcatctccagctgggcatctccagc 380
QY 286 caggctctccagccagggcatctccagccagggcatctccagctgggcatctccagc 345
DB 381 caggctctccagccagggcatctccagccagggcatctccagctgggcatctccagc 440
QY 346 cagagtgatctgtttagagcaaacacagctggggctgtaccatccgatactctcgc 405
DB 441 cagagtgatctgtttagagcaaacacagctggggctgtaccatccgatactctcgc 500
QY 406 caggtcagcacc-agcaaccagggccaccagggagagccaggtacg-agcgtgcccaag 463
DB 501 caggtcagcaccagcaaccagggccaccagggagagccaggtacgagccctgcccag 560
QY 464 ttcacctggcggggagggcc-agaagcagctacgcgtccatccggtgctgtctctccat 522
DB 561 ttcacctggcggggagggccaggaagcagctacgcgtccatccggtgctgtctctccat 620
QY 523 tgcctggtggtttcgtctcatctcctctccagttctgcaggggccacacagggatcag 582
DB 621 tgcctggtggtttcgtctcatctcctctccagttctgcagggttcgagggccacagggatnag 680
QY 583 gtacaaggagcagggagagagctgt-----cccaagcagctgtt-----cgctgtgacgggg 635
DB 681 gtacaaggagcagggagagagctgtcccaagcagctgttctcgtcgtgntgaacggggt 740

QY 636 tgggtgactgcaagctgaa 654
DB 741 ggggtgactgcaagctgga 759
RESULT 9
AAS91044
ID AAS91044 standard; cDNA; 1212 BP.
XX AAS91044;
AC AAS91044;
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26848.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG26857.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 26848; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1212 BP; 275 A; 348 C; 328 G; 261 T; 0 other;

Query Match 21.4%; Score 373.2; DB 23; Length 1212;
Best Local Similarity 83.6%; Pred. No. 7.1e-67;
Matches 475; Conservative 0; Mismatches 3; Indels 90; Gaps 1;
QY 804 gtgctaccggacaaccgaggttgcacacagggtttgcccacagcttttgcacacagcttctcaatttga 863

Db 404 gtgtcaccggacaacagaggttgcacacagggatttgcacacagcttctcaatttga 463
QY 864 gatacaactccacatccagaagaagctccacaggtgtgaatgcccttccacagcgtgata 923
Db 464 gatacaactccacatccagaagaagctccacaggtgtgaatgcccttccacagcgtgata 523
QY 924 tctccctccagtgttccactcgagactgagggccatgacgcggcgagatcggtggagggg 983
Db 524 tctctctccagtgttccactcgagactgagggccatgacgcggcgagatcggtggagggg 583
QY 984 cgtggtcctcggtatagcaagtgcccttggaagtgagtgacttgccacttgcgcaccaccaca 1043
Db 584 cgtggtcctcggtatagcaagtgcccttggaagtgagtgacttgccacttgcgcaccaccaca 643
QY 1044 tctgtgagggcagctcattgacccagtcgagtggtgtctcaactgcgcgccactgtttcteg 1103
Db 644 tctgtgagggcagctcattgacccagtcgagtggtgtctcaactgcgcgccactgtttcteg 703
QY 1104 t----- 1104
Db 704 ttaatgagtcagtgtagtactgtctgtctgaatttgcgtgaggtgaataataactagaga 763
QY 1105 -----gaccgggagaaggtctctggagggtctgga 1133
Db 764 tcttactagtaacagcccttgaccctgtctggaccgggagaaggtctctggagggtctgga 823
QY 1134 agtgtacgcgggcacacagcaactgcacagttgctgtaggcagcctccattgcccagaga 1193
Db 824 agtgtacgcgggcacacagcaactgcacagttgctgtaggcagcctccattgcccagaga 883
QY 1194 tcatcatcaacagcaattacacagatgagggagacgactatgacatcgccctcatgcggc 1253
Db 884 tcatcatcaacagcaattacacagatgagggagacgactatgacatcgccctcatgcggc 943
QY 1254 tgtcaagcccttgaccctgtccgctca 1281
Db 944 tgtccaagcccttgaccctgtccggtga 971

RESULT 10
AAS38533
ID AAS38533 standard; cDNA; 450 BP.
XX AC AAS38533;
AC AAS38533;
DT 17-DEC-2001 (first entry)
XX Novel human diagnostic and therapeutic gene #1591.
DE Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX Homo sapiens.
XX WO200166753-A2.
XX 13-SEP-2001.
XX 09-MAR-2001; 2001WO-US07787.
XX 09-MAR-2000; 2000US-0188609.
XX (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX WPI; 2001-530177/58.
DR
XX
PT New polynucleotides and polypeptides, useful for diagnosis and

PT treatment of breast, lung and colon cancer -
XX Claim 1; Page 994-995; 1193pp; English.
XX The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX Sequence 450 BP; 97 A; 166 C; 112 G; 75 T; 0 other;
SQ
Query Match 18.9%; Score 331.2; DB 22; Length 450;
Best Local Similarity 94.0%; Pred. No. 2.2e-58;
Matches 361; Conservative 0; Mismatches 8; Indels 15; Gaps 1;
QY 1 ctgagagaccatgagagggacagccagcaggaatgcatctccagcaagaacacattcagc 60
Db 65 ctgagagaccatgagagggacagccagcaggaatgcatctccagcaagaacacattcagc 124
QY 61 tggagcatctccagccagcagcattccagctgggacacctccagcggcgcatctccagc 120
Db 125 tggagcatctccagccagcagcattccagctgggacacctccagcggcgcatctccagc 184
QY 121 ccaggcatctccagccagcagcattccagctgggacacctccagcggcgcatctccagc 180
Db 185 ccaggcatctccagccagcagcattccagctgggacacctccagcggcgcatctccagc 244
QY 181 ccaggcatctccagctgggtacacctccagcggcgcatctccagcggcgcatctccagc 228
Db 245 ccaggcatctacagctgggtacacctccagcggcgcatctccagcggcgcatctccagc 304
QY 229 ---ggcatctccagccagcagcattccagccggcgcatctccggtctctggcatcttc 285
Db 305 ccaggcatctccagccagcagcattccagccggcgcatctccggtctctggcatcttc 364
QY 286 cagggtctccatccggcaggtcatctccgaggtccaggtcagctcgttgacacacccccaac 345
Db 365 cagggtctccatccggcaggtcatctccgaggtcagctcgttgacacacccccaac 424
QY 346 cagagtgatccttggtagagcaac 369
Db 425 cagagtgatccttggtagagcaac 448
RESULT 11
AAS79452
ID AAS79452 standard; cDNA; 705 BP.
XX AC AAS79452;
AC AAS79452;
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #15256.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PR

Db 1649 TCTCCCTGCATCTCTGTGCTGTGGGGAGAGCCTGAAGACCCCTCCCTGTGGTGGTGGGG 1590
QY 984 cgtggcctcggatagcaagtgcccttgcaagtgagctgcaacttcggcaccaccaca 1043
Db 1589 AGGAGCCTCTGTGGATCTTGGCCTTGGAGGTGAGCTAGCAGTCCAGTACGACAAACAGCAG 1530
QY 1044 tctgtgagagcagctcatagcagccagtgagtgctcactgcgcacactgtcttcteg 1103
Db 1529 TCTGTGGAGGGAGCATCTCTGGACCCCTACTGGTCTCAGGAGCCCTACTGCTTC--- 1472
QY 1104 tgaccgggagaggtccctggagggctgaagtgatgacaggggacacacacacacac 1163
Db 1473 --AGGAACATACCGATGTCTCACTGGAAGGTGGGGAGCTCAGACAACTGGGCA 1416
QY 1164 agtgctgagggcagcctcctcctgagagatcatcatcaacagcaattacacagtgagg 1223
Db 1415 GCTTCCC---ATCCCTGGCTGTGGCCAGATCATCATCTGAATTCACCCCATGTACC 1359
QY 1224 aggaagactatgacatgcctcctcctgagtgcttccaaagccctgacccctgtcgtccaa 1283
Db 1358 CCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTCACTTCTCAGGCACAG 1299
QY 1284 tcacacctgctgctccctcctcctgagagacacacctttagcctcaatgagacctgctga 1343
Db 1298 TCAGGCCCAATCTCTCTGCCCTCTCTTTGATGAGAGCTCACTCCAGCCACCCCACTCTGA 1239
QY 1344 tcacagcttggcagacagggagacagatgacagacatccctctcctccgggag 1403
Db 1238 TCATTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGCTGACATCTGCTCAGG 1179
QY 1404 tgacagtaactatcctcagctcctcctcctcctcctcctcctcctcctcctcctcctcct 1463
Db 1178 CGTGATCTCAGGTCTTATGACAGCAGACACGCTGCAATGCAGACGATGCTACAGGGGAAG 1119
QY 1464 tiaccacagagtgatgtctgagggacaccttcgtgagggcagagactcctccagggag 1523
Db 1118 TCACCGAAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACTGCCAGGGTG 1059
QY 1524 acagcgggggctctgtgtgagcagaaacacccctggtacctggcaggtgtcaca 1583
Db 1058 ACAGTGGTGGGCTCTGATGT---ACCAATCTGACCAAGTGGCATGTGGTGGCATGTGA 1002
QY 1584 gctggggcagagctgtggtcagagaaacacacctggtgtgtacacaaagtgacagag 1643
Db 1001 GCTGGGCTATGCTGCGGGGCCCCAGACCCCTCAGAGGATATACCAAGGTCTCAGCCT 942
QY 1644 ttcttccttgattacagcagagtgagcagagtgatgctgattcagaaatcctaacacc 1703
Db 941 ATCTCACTGGATCTACAATGTCTGGAAGCTGAGCTGTGAATGCTGCTGCCCTTTGCAG 882

RESULT 13
AA87154
ID AA87154 standard; cDNA; 2038 BP.
XX AC AA87154;
XX DT 27-SEP-1999 (first entry)
XX DE Human protease HUPM-6 cDNA.
XX KW Serine protease; human; HUPM-6; cell proliferation; cancer;
XX KW immune disorder; inflammation; therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 200..1507
XX FT /*tag= a
XX W09036550-A2.
XX PN
XX XX
XX PD

XX 12-JAN-1999; 99WO-US00655.
XX 16-JAN-1998; 98US-0008271.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
XX Tang YT, Yue H;
XX WPI: 1999-430616/36.
XX P-PSDB; AA06437.
XX Novel human protease molecules useful in the treatment of
XX developmental disorders and/or cancers
XX Claim 8; Page 86-87; 90pp; English.
XX This nucleotide sequence codes for HUPM-6 (see AA06437), a novel
XX human protease. HUPM-6 cDNA was initially identified in Incyte
XX Clone 1337018 from the colon cDNA library COLN013 using a
XX computer search for amino acid sequence alignments. The present
XX sequence is a consensus sequence derived from overlapping and/or
XX extended nucleic acid sequences: Incyte Clones 1271725 (TESFUT02),
XX 1337018, 586982 and 588598 (UTRSNOT01). A fragment comprising
XX nucleotides 900-949 of the present sequence can be used for
XX hybridisation. This sequence encompasses an active site residue.
XX Northern analysis shows expression of HUPM-6 in gastrointestinal,
XX and male and female reproductive cDNA libraries. Approximately 65%
XX of these libraries are associated with neoplastic disorders, and
XX 22% with the immune response. The invention provides 12 new human
XX proteases, i.e. HUPM-1 to -12 (see AA06432-43), and the
XX polynucleotides encoding them (see AA87149-60). Also provided are
XX vectors, host cells and methods for producing HUPM polypeptides, as
XX well as agonists and antagonists of HUPM. Methods for treating or
XX preventing cell proliferative disorders and immune disorders using
XX HUPM or HUPM antagonists are claimed.
XX Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 other;
Query Match 11.1%; Score 193.2; DB 20; Length 2038;
Best Local Similarity 55.8%; Pred. No. 3.3e-30;
Matches 435; Conservative 0; Mismatches 333; Indels 12; Gaps 3;
QY 924 tctccctcagtgctccactcgggactgagggccatgacggcgctggtggagggg 983
Db 759 tctccctcagtgcttctgctgtgggagagcctgaagacccctgtgtgtggggtggg 818
QY 984 cgtggcctcggatagcaagtgcccttggaagtgagtgagtgagtgagtgagtgagtgag 1043
Db 819 aggagggcctctgtggtattcttgccttgccagtgagtcagtcagtcagtcagtcagtcag 878
QY 1044 tctgtgagggcagctcattgacggccagtggtgtgctcactgcgcacactgtcttctg 1103
Db 879 tctgtgagggagcactcctggacccctcactggtgtcctcggcagccactgttc--- 934
QY 1104 tgacccgggagaggtcctggaggggctggaaggtgtacggggcagcagcagcactgcacc 1163
Db 935 --aggaaacataccagtggttcaactggaagtgcgggcagcagcagcagcagcagcagc 992
QY 1164 agtgctgagggcagcctcctcctgagtgatcatcatcaacagcaattacacagtgagg 1223
Db 993 gcttccc---atccctggctgtggcgaagatcatcatcatgaattcaacccactgtacc 1049
QY 1224 aggaagactatgacatgcctcctcctcctgagtgcttccaaagccctgacccctgtcgtccaa 1283
Db 1050 ccaagacaaatgacatgcctcctcctcctgagtgagtgagtgagtgagtgagtgagtgag 1109
QY 1284 tccacccctgctgctccctcctcctgagtgagtgagtgagtgagtgagtgagtgagtgag 1343
Db 1110 tcaggccactgtctgctccctcttctgtgagtgagtgagtgagtgagtgagtgagtgagtgag 1169

QY	1344	tcacaggctttgcaagaccaggagagacagatgacaaagacatcccccttctccggagg	1403	PR	16-SEP-1998;	98US-0100664.
Db	1170	ttattgatgggctttacgaagcagaatggagggaagatgtctgacatactgctgcagg	1229	PR	17-SEP-1998;	98US-0100683.
QY	1404	tgcaggtcaatctcatgactctcaagaatgcaatgactacttggtctatgacagttaac	1463	PR	17-SEP-1998;	98US-0100710.
Db	1230	cgtcagtcacaggctcatgacgacacacggtgcaatgcagacgatgcgtaccagggggaag	1289	PR	17-SEP-1998;	98US-0100711.
QY	1464	ttacccaagatgatgtgctggggacaccttgcgtggggcagagactcctgccaggag	1523	PR	17-SEP-1998;	98US-0100919.
Db	1290	tcaccgagaagatgatgtgctgaggcatcccggaagggtgtgacacctgccagggtg	1349	PR	17-SEP-1998;	98US-0100930.
QY	1524	acagcggggggcctctgtgtgagcagaacacccgtgttacctggcagggtgtcacca	1583	PR	18-SEP-1998;	98US-0100849.
Db	1350	acagtgtggcccttgatgt---accaatctgaccagtggtgtgggcacgttta	1406	PR	18-SEP-1998;	98US-0101014.
QY	1584	gctggggcacaggctgtgtggccagagaaacaaacctgtgtgtacaccaaagtacagaag	1643	PR	18-SEP-1998;	98US-0101014.
Db	1407	gctggggcgtatgctgcggggcccgagcaccgccagagagtatacaccagggtctcagcct	1466	PR	18-SEP-1998;	98US-0101014.
QY	1644	ttctcccttgatttacagcaagatggagcagcaggtgcgatttcagaaaaatcctaaccag	1703	PR	18-SEP-1998;	98US-0101014.
Db	1467	atctcaactgatactacaatgtctggaaggctgagtgttaatgctgctgccctttgcag	1526	PR	18-SEP-1998;	98US-0101014.
RESULT 14						
AAA37099						
ID	AAA37099 standard; cDNA; 2063 BP.					
AC	AAA37099;					
XX						
DT	08-AUG-2000 (first entry)					
DE	Human PRO1570 (UNQ776) cDNA sequence SEQ ID NO:274.					
KW	Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;					
KW	transmembrane; secretion; immunoadhesion; pharmaceutical; screening;					
OS	Homo sapiens.					
PN	WO200012708-A2.					
XX						
PD	09-MAR-2000.					
XX						
PF	01-SEP-1999; 99WO-US20111.					
XX						
PR	01-SEP-1998; 98US-0098716.					
PR	01-SEP-1998; 98US-0098749.					
PR	01-SEP-1998; 98US-0098750.					
PR	02-SEP-1998; 98US-0098803.					
PR	02-SEP-1998; 98US-0098821.					
PR	02-SEP-1998; 98US-0098843.					
PR	03-SEP-1998; 98US-0099536.					
PR	03-SEP-1998; 98US-0099596.					
PR	03-SEP-1998; 98US-0099598.					
PR	03-SEP-1998; 98US-0099602.					
PR	03-SEP-1998; 98US-0099642.					
PR	10-SEP-1998; 98US-0099741.					
PR	10-SEP-1998; 98US-0099754.					
PR	10-SEP-1998; 98US-0099763.					
PR	10-SEP-1998; 98US-0099792.					
PR	10-SEP-1998; 98US-0099808.					
PR	10-SEP-1998; 98US-0099812.					
PR	10-SEP-1998; 98US-0099815.					
PR	10-SEP-1998; 98US-0099816.					
PR	15-SEP-1998; 98US-0100385.					
PR	15-SEP-1998; 98US-0100388.					
PR	15-SEP-1998; 98US-0100390.					
PR	16-SEP-1998; 98US-0100584.					
PR	16-SEP-1998; 98US-0100627.					
PR	16-SEP-1998; 98US-0100661.					
PR	16-SEP-1998; 98US-0100662.					
PR	03-NOV-1998; 98US-0100656.					


```

PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 03-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH ) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2001-602746/88.
DR P-PSDB; AAU29188.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 2; Fig 329; 774pp; English.
XX
XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can
CC be used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

```

```

Query Match 11.1%; Score 193.2; DB 22; Length 2063;
Best Local Similarity 55.8%; Pred. No. 3.3e-30;
Matches 435; Conservative 0; Mismatches 333; Indels 12; Gaps 3;
QY 924 tctctccactgttccactgcggactgagggccatgaccggcgagatcgtgaggggg 983
D 760 tctctccactgtcttgcctgtggaagagcctgaagaccccccggtgtgtgtggggg 819
QY 984 cgtcggcctcgatagacaagtgccttggcaagtgcacttcgtggcaccaccaca 1043
D 820 agaggcctctgtgattcttgccttggcaggttcagcatccagtagacgaacacgc 879
QY 1044 tctgtgagcagcgtcattgacgccagtggtgtcactgcgccactgcttcttgc 1103

```

```

Db 880 tctgtgaggaggacatcctgtgaaccccccaactgggtctctcagcgagcccaactgttc--- 935
QY 1104 tgaccggggagagggtcctgtgagggtgtaacggtgtacgcggcgaccagcaacctgcacc 1163
Db 936 --aggaaacataccgatgtgttcaactggaaggtgcggcgaggtctcagacaaactgggca 993
QY 1164 agttccttgcaggcgcctccattgcggagatcatcatcaacagacaattacacgatgagg 1223
Db 994 gcttccc--atccctggtgtggtgccaagatcatcatcattgaattcaaccccatgtacc 1050
QY 1224 aggaagactatgacatgcctccatgcggtgttccaagccctgacccctgccttcgcctaca 1283
Db 1051 ccaagacaatgacatgcctccatgaagtcagttcccaactcacttcttcaggccacag 1110
QY 1284 tccacctgttgcctcccatgcagacagacacctttagcctcaatgagacctgctgga 1343
Db 1111 tcaggcccatctgtcgccttcttggatgagagctcactccagccacccactctgga 1170
QY 1344 tcacaggcttggcaagaccaggagacagatgacaaagacatcccccttctccggagg 1403
Db 1171 tcattggtgggttttacgaagcagaatggagggaagatgtctgacatactgtcgagg 1230
QY 1404 tgcaggtcaatctcatcgacttcaagaaatgcaatgactacttggctctatgacagttacc 1463
Db 1231 cgtcagtcacagggtcattgacagcacacgggtgcaatgacagcagatgcgtaccagggggaag 1290
QY 1464 ttaccccaagatgattgtgtggtgggacaccttctgtggggcagagactcctgccaggag 1523
Db 1291 tcaccgagaagatgattgtgcaggcatcccggaagggtgtggacacctgccagggtg 1350
QY 1524 acagcggggggcctcttctgtgtgagcagaaacacccgctggtacctggcagggtgtcacca 1583
Db 1351 acagtgtgtggccctgatgt--accaatctgaccagtgcatgtgtgtgggcatcgta 1407
QY 1584 gctggggcacagcgtgtggccagagaaacaaacctggtgtgtacaccaaagtgcacagaag 1643
Db 1408 gctggggctatggtcgcggggcccgagaccaccagagagatatacacaaggcttcagcct 1467
QY 1644 ttcttcctggatttacagcaagatggagagcagggtgcgatttcagaaaaatcctaaccag 1703
Db 1468 atctcaactggatctacaatgtctggaagggtgagctgtaagtgtgctgctgcccctttgcag 1527

```

Search completed: August 19, 2002, 18:43:44
Job time: 5987 sec